DOCUMENT RESUME

ED 116 448 FL 005 227

AUTHOR Lea, Wayne A.
TITLE Syntactic Boundaries and Stress Patterns in Spoken English Texts. Univac Report No. PX 10146.
INSTITUTION Sperry Univac, St. Paul, Minn. Defense Systems Div.
REPORT NO PX-10146
PUB DATE 31 Mar 73
NOTE 117p.

EDRS PRICE MF-$0.76 HC-$5.70 Plus Postage
DESCRIPTORS Acoustic Phonetics; Algorithms; Articulation (Speech); Auditory Perception; *Computational Linguistics; Computer Programs; *Intonation; Pattern Recognition; Phonetic Analysis; Phonological Units; Phonology; Sentences; Speech; *Stress (Phonology); Suprasegmentals; *Syllables; *Syntax
IDENTIFIERS Frequency Contours; Speech Recognition; Syntactic Boundaries

ABSTRACT

This report covers research conducted between July 1972 and March 1973. Experiments were conducted on the automatic detection of constituent boundaries and location of stressed syllables by analysis of fundamental frequency and energy contours, for recordings of six talkers reading the Rainbow Script, two talkers reading a paragraph composed of monosyllabic words, and ten talkers involved in speaking sentences pertinent to man-computer interaction. A program was implemented which successfully detects over 80 percent of all boundaries between major syntactic constituents, by the use of fall-rise valleys in fundamental frequency contours. A panel of three listeners provided judgments of which syllables were stressed, unstressed, or reduced in the speech texts. Questions yielded more stress level confusions than declaratives or commands. An algorithm was devised for locating stressed syllables as high energy portions of speech with rising or nonfalling fundamental frequency. This algorithm succeeded in locating 85 percent of all syllables that had been perceived as stressed by two or more listeners. Further work will involve implementation of the stressed syllable location algorithm, refinements of syntactic boundary predictions and detection procedures, further tests with designed speech texts, and applications to distinctive features estimation and syntactic parsing. (Author/KM)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED116448

SPERRY UNIVAC
DEFENSE SYSTEMS

SYNTACTIC BOUNDARIES AND STRESS PATTERNS IN SPOKEN ENGLISH TEXTS

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

by

Wayne A. Lea

Defense Systems Division
Univac Park
St. Paul, Minnesota 55165

March 31, 1973

Document No.
PX 10146

FL 005 227

UNIVAC

1. DOCUMENT NO.	2. GOVERNMENT ACCESSION NO.	3. RECIPIENT'S CATALOG NO.
PX 10146	N/A	N/A

4. TITLE AND SUBTITLE	5. REPORT DATE
Syntactic Boundaries and Stress Patterns in Spoken English Texts	March 31, 1973
	6. PERFORMING ORGANIZATION CODE 6670
7. AUTHOR (S) Wayne A. Lea	8. PERFORMING ORGANIZATION REPORT NO.
9. PERFORMING ORGANIZATION NAME AND ADDRESS Sperry Univac Defense Systems Division Speech Communications Group P. O. Box 3525 St. Paul, Minnesota 55165	10. WORK UNIT NO.
	11. CONTRACT OR GRANT NO.
12. SPONSORING AGENCY NAME AND ADDRESS	13. TYPE OF REPORT AND PERIOD COVERED Speech Processing Research July, 1972-March, 1973
	14. SPONSORING AGENCY CODE

15. SUPPLEMENTARY NOTES

16. ABSTRACT

Experiments were conducted on the automatic detection of constituent boundaries and location of stressed syllables by analysis of fundamental frequency and energy contours, for recordings of six talkers reading the Rainbow Script, two talkers reading a paragraph composed of monosyllabic words, and ten talkers involved in speaking sentences pertinent to man-computer interaction. A program was implemented which successfully detects over 80% of all boundaries between major syntactic constituents, by the use of fall-rise valleys in fundamental frequency contours. A panel of three listeners provided judgments of which syllables were stressed, unstressed, or reduced in the speech texts. Judgments from two listeners were quite consistent from time to time, and the two listeners particulary agreed with each other as to which syllables were stressed. The third listener gave less consistent results. Stress judgments based on the written text alone (without hearing the speech) were about as consistent from time to time or listener to listener as were results with speech, but the "NO SPEECH" judgments were different from the "SPEECH"-determined judgments, particulary for spontaneous utterances. Questions yielded more stress level confusions than declaratives or commands. An algorithm was devised for locating stressed syllables as high energy portions of speech with rising or non-falling fundamental frequency. This algorithm succeeded in locating 85% of all syllables that had been perceived as stressed by two or more listeners. Further work will involve implementation of the stressed syllable location algorithm, refinements of syntactic boundary predictions and detection procedures, further tests with designed speech texts, and applications to distinctive features estimation and syntactic parsing.

17. KEY WORDS	18. DISTRIBUTION STATEMENT
Syntactic Boundary Detection Linguistic Stress Intonation Prosodic Features Speech Recognition	See Attached List

19. SECURITY CLASSIF. (OF THIS REPORT)	20. SECURITY CLASSIF. (OF THIS PAGE)	21. NO. OF PAGES	22. PRICE
Unclassified	Unclassified	x+107	

DISTRIBUTION LIST

ATIC File

W. J. Malloy

G. M. Workman

C. F. Mittelstadt

M. F. Medress

T. E. Skinner

C. W. Glewwe

ACKNOWLEDGEMENTS

Mark Medress and Toby Skinner of the Sperry Univac Speech Communications Group participated in the stress perception experiments reported herein, and provided the acoustic analyses of the speech. We are indebted to George W. Hughes and Kung-Pu Li of Purdue University for providing the recordings of the Rainbow Script and the Monosyllabic Script, and for developing the basic perceptual testing procedures which have been modified for use in this study. The ARPA Sentences were originally recorded by five ARPA contractors, and are presently being studied in detail within the ARPA Speech Understanding Research Program.

TABLE OF CONTENTS

ABSTRACT		ii
DISTRIBUTION LIST		iii
ACKNOWLEDGEMENTS		iv
LIST OF FIGURES		vii
LIST OF TABLES		x
1. INTRODUCTION		1
2. SELECTED SPEECH TEXTS		4
3. CONSTITUENT BOUNDARY DETECTION		8
	3.1 Obtaining F_0 and Energy Measurements	8
	3.2 The Constituent Boundary Detector	8
	3.3 Boundaries Detected in the Rainbow Script	10
	3.4 Boundaries Detected in the Monosyllabic Script	14
	3.5 Boundaries Detected in the ARPA Sentences	16
	3.6 Summary of Boundary Detection Results	17
4. PERCEIVED STRESS PATTERNS		20
	4.1 Experimental Methods	20
	4.2 Majority Judgments of Stress Levels	22
	4.3 Effects of the Individual Listener on Stress Perceptions	24
	4.4 Consistency of Perceptions From Time to Time	29
	4.5 Comparing Stress Judgments with Speech to Those Without Speech	34
	4.6 Effects of Sentence Type on Stress Judgments	40
	4.7 General Conclusions About Stress Perceptions	42
5. STRESSED SYLLABLE LOCATION FROM ACOUSTIC DATA		46
	5.1 Correlates of Stress in F_0 Contours	46
	5.2 Energy-Integral Cues to Stress	51
	5.3 An Algorithm for Stressed Syllable Location	54
	5.3.1 Finding the First Stressed Syllable in a Constituent	54
	5.3.2 Finding Other Stressed Syllables in a Constituent	58
	5.4 Comparison of Algorithmic Locations with Perceived Stress Patterns	59

6. CONCLUSIONS AND FURTHER WORK 64

7. REFERENCES 68

APPENDIX A. CONSTITUENT BOUNDARY DETECTION RESULTS 71

APPENDIX B. DETAILS OF PERCEIVED STRESS PATTERNS 79

APPENDIX C. STRESSED SYLLABLES LOCATED BY ALGORITHM 95

LIST OF FIGURES

1. Boundary Detection Results for a Portion of the Rainbow Script Spoken by Six Talkers 11

2. Summary of Stress Judgments by Three Listeners, for Talker ASH Reading the Rainbow Script 23

3. Percentages of Stress Judgments that Differ from One Listener to Another, for Each Speech Text, and With Each Speaker and the NO SPEECH Conditions 26

4. Percentages of Listener-to-Listener Confusions in Assigned Stress Levels for Each Text and Talker, with Unstressed-Reduced, Stressed-Unstressed, and Stressed-Reduced Confusions Separately Graphed 27

5. Percentages of Stress Judgments that Differ from One Trial to Another, for Each Speech Text, and With Each Speaker and the NO SPEECH Conditions 31

6. Percentages of Repetition-to-Repetition Confusions in Assigned Stress Levels by Each Listener, for Each Text and Talker, with Unstressed-Reduced, Stressed-Unstressed, and Stressed-Reduced Confusions Separately Graphed 32

7. Percentages of Confusions in Assigned Stress Levels for NO-SPEECH versus SPEECH Conditions, for Each Text and Talker 38

8. Effects of Individual Sentence Type on the Percentages of Stress Level Confusions for the ARPA Sentences 41

9. Tune I and Tune II Intonation Contours 48

10. Each Major Constituent of a Sentence is Assumed to Exhibit a Rapidly-Rising, Gradually-Falling "Archetype Constituent Contour", Riding on the Overall Tune I Contour of the Sentence 50

11. Increases in F_o, Above the Archetype Contour for a Constituent, Are Assumed to be Associated with Stressed Syllables 52

12. Computer Printout of the Fundamental Frequency and Broadband Speech Energy Functions for Each 10 ms of the Question "Who is the owner of utterance eight?" 55

A-1. Complete Boundary Detection Results for the Rainbow Script Spoken by Six Talkers 72

A-2. Complete Boundary Detection Results for the Monosyllabic Script for Talkers ASH and GWH 73

A-3. Complete Boundary Detection Results for the 13 ARPA Sentences 74

A-4. Effects of Threshold Size on Boundary Detection Results, for the 6ARPA Sentences 77

B-1. Sample of the Sheets Used for Marking Stress Judgments (Listener MFM) 80

B-2. Summary of Stress Judgments by Three Listeners, for Talker ASH Reading the Rainbow Script 81

B-3. Summary of Stress Judgments by Three Listeners, for Talker GWH Reading the Rainbow Script 82

B-4. Summary of Stress Judgments by Three Listeners, for Talker WB Reading the Rainbow Script 83

B-5. Summary of Stress Judgments by Three Listeners, for Talker JP Reading the Rainbow Script 84

B-6. Summary of Stress Judgments by Three Listeners, for Talker PB Reading the Rainbow Script 85

B-7. Summary of Stress Judgments by Three Listeners, for Talker ER Reading the Rainbow Script 86

B-8. Summary of Stress Judgments by Three "Listeners", When Given Only the Written Text of the Rainbow Script (NO SPEECH) 87

B-9. Summary of Stress Judgments by Three Listeners, for Talker ASH Reading the Monosyllabic Script 88

B-10. Summary of Stress Judgments by Three Listeners, for Talker GWH Reading the Monosyllabic Script 89

B-11. Summary of Stress Judgments by Three "Listeners", When Given Only the Written Text of the Monosyllabic Script (NO SPEECH) 90

B-12. Summary of Stress Judgments by Three Listeners, for the 6ARPA Sentences as Spoken 91

B-13. Summary of Stress Judgments by Three "Listeners", When Given Only the Written Text of the 6ARPA Sentences (NO SPEECH) 92

B-14. Summary of Stress Judgments by Three Listeners, for the 7ARPA Sentences as Spoken 93

B-15. Summary of Stress Judgments by Three "Listeners", When Given Only the Written Text of the 7ARPA Sentences (NO SPEECH) 94

C-1. Flowchart of the Algorithm for Locating Stressed Syllables 96

C-2. Comparison of Algorithmically Located Stressed "Syllables" with Perceived Stress Patterns, for Talker ASH Reading the Rainbow Script 97

C-3. Comparison of Algorithmically Located Stressed "Syllables" with Perceived Stress Patterns, for Talker GWH Reading the Rainbow Script 98

C-4. Comparison of Algorithmically Located Stressed "Syllables" with Perceived Stress Patterns, for Talker WB Reading the Rainbow Script 99

C-5. Comparison of Algorithmically Located Stressed "Syllables" with Perceived Stress Patterns, for Talker JP Reading the Rainbow Script 100

C-6. Comparison of Algorithmically Located Stressed "Syllables" with Perceived Stress Patterns, for Talker PB Reading the Rainbow Script 101

C-7. Comparison of Algorithmically Located Stressed "Syllables" with Perceived Stress Patterns, for Talker ER Reading the Rainbow Script 102

C-8. Comparison of Algorithmically Located Stressed "Syllables" with Perceived Stress Patterns, for Talker ASH Reading the Monosyllabic Script 103

C-9. Comparison of Algorithmically Located Stressed "Syllables" with Perceived Stress Patterns, for Talker GWH Reading the Monosyllabic Script 104

C-10. Comparison of Algorithmically Located Stressed "Syllables" with Perceived Stress Patterns, for the 6ARPA Sentences 105

C-11. Comparison of Algorithmically Located Stressed "Syllables" with Perceived Stress Patterns, for the 7ARPA Sentences 106

LIST OF TABLES

I. Boundary Detection Results for Rainbow Script 13

II. Boundary Detection Results for Monosyllabic Scrip 15

III. Summary of Boundary Detection Scores 18

IV. Stressed Syllable Location Scores 62

A-I. Boundary Detection Results for Various Sentence Types 75

1. INTRODUCTION

Computers that understand speech are expected to facilitate natural man-machine interaction, but the problems involved demand the attention of several disciplines, including linguistics, computer systems design, perception theory, speech research, and engineering. Linguistic and perceptual arguments, in particular, suggest that devices which recognize speech will have to make use of grammatical structure ("syntax") in early stages of the recognition procedures (Lea, 1972a,b; 1973b; Lea, Medress, and Skinner, 1972a). This can be accomplished, in part, by using certain acoustic correlates of prosody, such as energy and voice fundamental frequency contours, to segment the speech into grammatical phrases, and to identify those syllables that are given prominence, or stress, in the sentence structure.

In this paper, methods are described for (1) detecting syntactic boundaries from fall-rise patterns in voice fundamental frequency (F_0) contours, then (2) locating stressed syllables, within each syntactic unit, as high-energy portions of the speech which exhibit significantly high and rising (or, in some cases, non-falling) F_0 contours. The algorithmic locations of stressed syllables are compared with listeners' perceptions of stress, to determine how well the algorithmic results correspond with perceived prominence.

Once the connected speech is segmented into phrases, and stressed syllables are located, the Univac speech recognition strategy would call for a partial distinctive features analysis within each stressed syllable. Consonants and vowels are expected to be more clearly articulated and easier to distinguish in stressed syllables, than in unstressed or reduced syllables (cf. Lea, Medress, and Skinner, 1972b), where articulation (and consequent acoustic information) is not as precise or consistent from talker to talker or time to time.

Next, the partial distinctive features description would be matched with generated or stored patterns for possible stressed syllables or words in the lexicon. Then a guess as to the word content of the constituent would be made, based on the reliable feature information from the stressed syllables, plus other reliable data within the constituent (such as presence of coronal strident

fricatives, etc.; cf. Medress, 1972). Each guess as to constituent identity would be combined with those for other constituents in the sentence until a satisfactory set of hypotheses for all constituents yielded the grammatical, meaningful sentence.

In addition to aiding partial distinctive features estimation, the presence of syntactic boundaries and the positions of stressed syllables are expected to help guide syntactic parsers (Lea, 1972a). For example, an investigation has begun of the feasibility of using prosodically-detected syntactic boundaries to affect the priority order on transition arcs and the pop-up procedures in parsers based on Wood's transition network grammar (Woods, 1971).

In the remainder of this report, the encouraging successes in applying the boundary detector and a stressed syllable location algorithm will be presented. In section 2, the speech texts selected for this research are given, and their relative merits for prosodic analyses are outlined. Then, in section 3, an algorithm is described for detecting constituent boundaries from fall-rise patterns in F_o contours, and its application to the selected texts is shown to provide successful detection of over 80% of all predicted syntactic boundaries.

In section 4, experiments are described which show that several listeners rather consistently classified all syllables in the spoken texts into one of three categories - stressed, unstressed, or reduced. Issues of interest with regard to these stress perceptions are the effects of individual talkers, individual listeners, various texts, how consistent the listener's perceptions are from time to time, whether the listener can predict stress levels given only the written text (without listening to the speech recordings), and which stress levels (stressed, unstressed, or reduced) are most consistently assigned. The majority decisions of the team of listeners provide the standard by which a stressed syllable algorithm can be judged.

In section 5, an algorithm is described for locating stressed syllables within the phrases delimited by the constituent boundary detector. This algorithm, which is based on previous intonation theories and studies of acoustic correlates of stress, assumes that stressed syllables will be accompanied by rising or non-falling F_o and large energy integral. The results show that about 85% of all syllables that were usually judged as stressed by a majority of the listeners were also found by the algorithm.

In section 6, further work is outlined, to improve the algorithms for syntactic segmentation and stressed syllable location, and to combine partial distinctive features analysis within the stressed syllables with aids to syntactic parsing. Such efforts would yield critical portions of the proposed speech recognition strategy.

Appendices are included to detail the results in constituent boundary location (Appendix A), perceived stress patterns (Appendix B), and the results of algorithmic location of stressed syllables (Appendix C).

2. SELECTED SPEECH TEXTS

To test the algorithms for boundary detection and stressed syllable location, speech texts had to be chosen, recorded, submitted to listeners for stress perceptions, and analyzed by the computer programs. The primary text chosen for these studies was the first paragraph of the "Rainbow Passage" (Fairbanks, 1940). It reads as follows:

> "When the sunlight strikes raindrops in the air, they act like a prism and form a rainbow. The rainbow is a division of white light into many beautiful colors. These take the shape of a long round arch, with its path high above, and its two ends apparently beyond the horizon. There is, according to legend, a boiling pot of gold at one end. People look, but no one ever finds it. When a man looks for something beyond his reach, his friends say he is looking for the pot of gold at the end of the rainbow."

This text (hereinafter called the Rainbow Script) has been used extensively in studies of prosodic patterns in speech, and has the advantage of being a well-known semantically-connected text of declarative sentences, with a variety of grammatical phrase structures (cf. Lea, Medress, and Skinner, 1972a). It was recorded by six talkers (four male, two female) in a quiet room at Purdue University.

In texts like the Rainbow Script, the factors determining positions of stress within words (lexical stress) are compounded with sentence structure effects on stress (cf. Chomsky and Halle, 1968; Halle and Keyser, 1971). Another text which was composed of only monosyllabic words was also analyzed, to eliminate or minimize lexical stress effects. This text, read by two of the six talkers who had read the Rainbow Script, is the first paragraph of a short story:

> "John and I went up to the farm in June. The sun shone all day, and wind waved the grass in wide fields that ran by the road. Most birds had left on their trek south, but old friends were there to greet us. Piles of wood had been stacked by the door, left there by the man who lives twelve miles down the road. The stove would not last till dawn on what he had cut, so I went and chopped more till the sun set."

Lea (1972a,b) had previously processed recordings of this text (hereinafter referred to as the Monosyllabic Script) for constituent boundary detection at Purdue University. Comparing his previous results with the boundary detections found by the Univac implementation of his algorithm helped verify the new algorithm.

Both the Rainbow Script and the Monosyllabic Script involve read speech, all of declarative structure. To evaluate the boundary detection and stressed syllable location techniques with questions, commands, and declaratives of direct utility in man-machine interactions, thirteen sentences were selected from actual recordings by five contractors who are developing speech understanding systems for the Advanced Research Projects Agency (ARPA) of the Department of Defense (cf. Newall, et al., 1971). Most of these sentences were not read, but were composed on the spot in simulated protocols of man-machine interaction. The semantic context of each sentence was pertinent to a particular task domain adopted by the builder of a speech understanding system, such as retrieving information on lunar rock samples (Woods, 1971), other information-retrieval tasks, instructing a robot to move objects in a block world (Walker, 1973), or voice programming.

These thirteen sentences are as follows:

1. (LS21) Who's the owner of utterance eight?
2. (LM13) Display the phonemic labels above the spectrogram.
3. (B27) Do any samples contain troilite?
4. (B10) What is the average uranium lead ratio for the lunar samples?
5. (RB6) Do you have any right square boxes left?
6. (RB16) Put the other red block on the red block.
7. (LM3) Who is the owner of utterance eight?
8. (B35) Do any samples contain tridymite?
9. (RA19) Would you move the stack of right circular cylinders to the right by half a square?
10. (RC8) Place the red triangle two squares back from the front of the floor in the middle.

11. (CV1300) Alpha becomes alpha minus beta.
12. (CV2300) Alpha gets alpha minus beta.
13. (D10) Repeat where key work equals Gauss elimination or key word equals eigenvalues.

The recordings of these 13 "ARPA Sentences" involved ten different talkers, each one saying one or more of the sentences, as indicated by the distinguishing alphabetic code for each talker, shown within the parentheses.[1]

We shall distinguish the first six ARPA sentences (hereinafter referred to as "6ARPA Sentences") from the last seven (referred to as "7ARPA Sentences"), since the first six are being studied extensively by various ARPA contractors, while the seven additional sentences were selected by the author to provide several additional interesting syntactic constructions and many more syntactic boundaries than the first six had provided. These sentences show a variety of sentence types (three questions with interrogative words (who, what), three yes-no questions, four imperatives, one "polite" command or request, and two declarations), with emphasis on questions and commands, which are expected to be of major interest to man-computer communications. Some of the structures (as in D10) are not usual English forms, but obey syntax equations being designed into the restricted parsers of speech-understanding systems. Yet, each sentence has at least one interesting phrase structure or contrast with another possible structure, such as the compound nouns in D10, sequence of noun phrases and prepositional phrases in RA19 and RC8, or the "minimal pair" contrasts between B27 and B35 or LS21 and LM3.

1. The first letter of the code identifying each sentence, as shown within the parentheses of this list, indicates the ARPA contractor which recorded the sentence (B = Bolt, Beranek, and Newman; C = Carnegie Mellon University; D = Systems Development Corporation; L = Lincoln Laboratories; and R = Stanford Research Institute). The second letter, when it appears, identifies which talker from that organization spoke the associated sentence, or, in the case of CV codes, it marks the task as voice programming. Numbers in the code indicate the order in which the sentence appeared in that organization's protocol of utterances. This complex code is included here since these same utterances are being studied, under such identifiers, by various ARPA contractors.

The recordings of all these speech texts provide a total of 3[illegible] predicted syntactic boundaries and 1128 syllables for evaluating the effectiveness of the boundary detection and stressed syllable location algorithms.

3. CONSTITUENT BOUNDARY DETECTION

3.1 Obtaining F_o and Energy Measurements

The speech recordings for the Rainbow Script, Monosyllabic Script, and ARPA Sentences were digitized and submitted to computer programs that obtained fundamental frequency and broadband (5 KHz) energy measures for each 10 milliseconds (ms) of speech. The fundamental frequency measure in Hertz, as provided by autocorrelating the center-clipped waveform (Sondhi, 1958), was also converted to eighth-tones, yielding a log frequency scale for relative measurements. The energy measure was obtained, using a 25.6 ms Hanning window, from the sum of the squares of the time waveform values (Blackman and Tukey, 1958), followed by a conversion to a relative (dB) scale. Smoothed spectra from a linear prediction scheme (Makhoul, 1972) and formant tracks were also obtained, but were not used for the present studies except to help determine where in the text each F_o or energy effect occurred.

The F_o and energy measurements were plotted versus time by a computer plotting program. For examples of F_o and energy plots, see (Lea, Medress, and Skinner, 1972a, p. 25) or Figure 12 in section 5 of this report.

3.2 The Constituent Boundary Detector

The F_o measurements were then submitted to an algorithm for detecting boundaries between major grammatical constituents. This boundary-detection algorithm (Lea, 1972a,b; 1973b) is based on an assumption that F_o will usually decrease (about 7% or more) at the end of each major syntactic constituent, and then increase (about 7% or more) either at the beginning of the following constituent or after any unstressed syllables at the beginning of that following constituent. Experimenting with fundamental frequency contours in over 500 seconds of speech (including short stories, newscasts, weather reports, and excerpts from conversations, spoken by nine talkers), Lea had shown that over 80% of all syntactically predicted boundaries were correctly detected (Lea, 1972a,b; 1973b). Lea had, however, observed that about half of all "missing" boundaries were due to predicted boundaries between noun phrases and following

auxiliary verbs or main verbs. He concluded that such noun phrase-verbal boundaries should not always be expected in phonological structure.

Detecting such syntactic structure from F_o contours is complicated by the fact that, at consonant-vowel (and vowel-consonant) boundaries, variations in F_o occur which may be confused with the changes marking syntactic boundaries. False (syntactically unrelated) boundary detections resulted from F_o variations at these boundaries between vowels and consonants, but most such false alarms could be eliminated by setting a minimum percent variation (about 10%) in F_o for a boundary detection. A detailed study of F_o variations at phonetic boundaries (Lea, 1972a, Ch. 4; cf. also Lea, 1973a) clearly indicated that such phonetically-dictated changes in F_o would rarely exceed about 10%.

The boundary detection algorithm also detects clause and sentence boundaries wherever long (350 ms) stretches of unvoicing (i.e., "pauses") occur.

While several improvements could be made in the original algorithm, and in the previous predictions as to where syntactic boundaries should be detected, the present studies were done with substantially the same algorithm, implemented as a FORTRAN program at the Univac Speech Communications Laboratory. One exception is that the results to be reported for the Rainbow Text were obtained by a hand analysis, strictly following Lea's algorithm, but including one refinement which eliminated some false boundaries resulting from large (7% or greater) variations in F_o that only last for one 10 ms time sample. The original hypothesis that boundaries would occur between noun phrases and verbals was also maintained, until a precise formulation of when it fails could be established. As Lea had previously suggested (Lea, 1972b), boundaries were not predicted between pronouns and following verbals.

3.3 Boundaries Detected in the Rainbow Text

Figure 1 shows some typical boundary marking for a portion of the Rainbow Script, spoken by male talkers ASH, GWH, WB, and JP, and female talkers PB and ER. Detected boundaries that corresponded to predicted boundaries are shown by vertical bars below the place in the speech where they occurred. Unpredicted F_o valleys which could be correlated with lower level syntactic boundaries are shown by columns of dots. False boundaries, due to nonsyntactic effects such as F_o changes at consonant-vowel boundaries, are shown by question marks. When a predicted boundary was missing from the detection, an asterisk is marked at the predicted position for the syntactic boundary. Sentence boundaries, determined by "pauses" of long-term unvoicing, are marked by dollar signs ("S's" with the vertical bars of "predicted" boundaries).

Thus, predicted boundaries are shown to be detected for all talkers between the copulative is and the object noun phrase a division, and before prepositional phrases of white light and into many beautiful colors. The predicted boundary between the noun-phrase subject The rainbow and the verbal is was detected for five of the talkers, but missed in the F_o contours of talker PB. These noun phrase/verbal boundaries are more frequently missed in other instances in the texts, as may be seen from Figures A-1, A-2, and A-3 of Appendix A, which illustrate the complete set of boundary detection results for all the texts and talkers.

Sometimes the rise in F_o into a constituent may be delayed due to initial weakly stressed syllables or function words like a, of, into, etc. The bottom of the F_o valley may then be delayed until within such weak beginnings of constituents, such as illustrated by the horizontal arrows in the beginnings of constituents such as a division, of white light, and into many beautiful colors. This delay is considered a predictable result of the stress patterns, and such displaced boundaries are still considered correctly detected. These delays, however, illustrate that the algorithm is not locating syntactic boundaries, only detecting them.

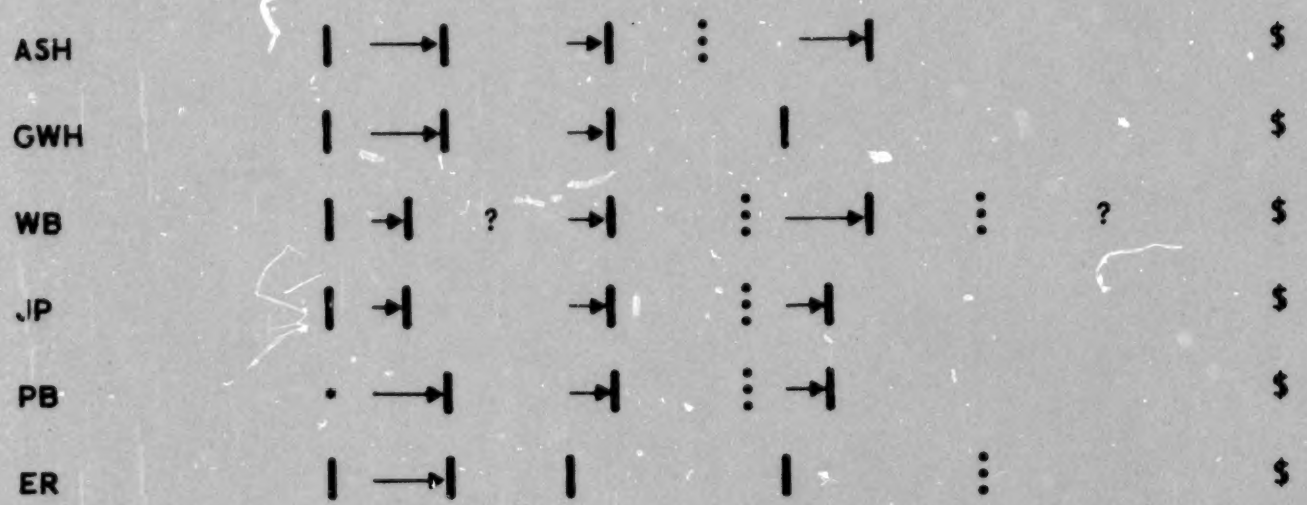


Figure 1. Boundary Detection Results for a Portion of the Rainbow Script Spoken by Six Talkers. Predicted boundaries that were detected are shown by vertical bars (|); boundaries between minor syntactic constituents (that were detected but not predicted) are shown by columns of dots (⋮); and false (syntactically unrelated) boundaries that were detected are shown by question marks (?). Asterisks (*) are marked at positions where syntactic boundaries had been predicted but had not been detected by the boundary detection algorithm.

Table I summarizes the boundary detection results obtained from the hand analysis of the Rainbow Text, as spoken by the six talkers. Forty-two constituent boundaries had been predicted by the independent syntactic analysis based on an intuitive constituent-structure division of the sentences, and previous experience with fundamental frequency patterns. Table I shows that the number of correctly detected boundaries (second column from the left) varied somewhat from talker to talker, yielding detection scores (third column from the left) that ranged from 67% to 86% of all predicted boundaries that were detected. The average detection score (79%) is very close to the 81% scores obtained by Lea in previous experiments with other texts (Lea, 1973b).

Also tabulated in Table I are the numbers of "extra" detected boundaries (fourth column from the left) that related to boundaries between minor syntactic constituents, but which had not been predicted by the particular syntactic analysis used. An improved procedure for predicting prosodically-marked syntactic boundaries might include these among the "predicted" boundaries for future studies. The last column in Table I shows the number of false (syntactically unrelated) boundaries that were found in each spoken text. These "false alarms" are considerably reduced in number from Lea's previous results (1972b, p. 66), because of the refinement that requires F_o maxima and minima to last for at least two time segments (20 ms).

All boundaries between matrix sentences (five per talker) were accompanied by long (350 ms or more) durations of unvoicing, and were thus correctly marked as sentence boundaries. However, boundaries between embedded sentences (that is, clause boundaries within matrix sentences), while always marked as constituent boundaries by F_o valleys, were accompanied by pauses in only 14 of the expected 24 instances.

An apparent "sentential pause" that had <u>not</u> been predicted (but which is not surprising) occurred after the parenthetical phrase <u>according to legend</u>, for two of the six talkers. No pauses of 350 ms or longer occurred at other than such major syntactic boundaries.

TABLE I.

BOUNDARY DETECTION RESULTS FOR RAINBOW SCRIPT

TALKER	NUMBER OF PREDICTED BOUNDARIES CORRECTLY DETECTED	% OF ALL PREDICTED BOUNDARIES CORRECTLY DETECTED	NUMBER OF "EXTRA" BOUNDARIES DETECTED	NUMBER OF "FALSE" BOUNDARIES DETECTED
ASH	30	72%	2	0
GWH	35	83%	4	1
WB	34	81%	7	6
JP	36	86%	8	2
PB	28	67%	2	0
ER	35	83%	6	0
AVERAGE ALL TALKERS	33	79%	5	2

These results for the Rainbow Script are very similar to those found earlier by Lea for another set of talkers reading weather reports, newscasts, and other texts, and for short conversational excerpts.

3.4 Boundaries Detected in the Monosyllabic Script

Figure A-2 in Appendix A shows the complete set of boundary detections for the Monosyllabic Script, as found by the Univac implementation of Lea's original algorithm. These computer-derived results are similar to those shown in Figure 1 for the Rainbow Script, and agree substantially with results reported by Lea (1972b, p. 199) for two other talkers.

Table II summarizes the boundary detection results for the Monosyllabic Script. The scores of 86% (for ASH) and 80% (for GWH) show substantial improvement from the respective scores of 73% and 66% correct detection of predicted boundaries reported for the same two talkers (ASH and GWH) in Lea's earlier hand analysis (Lea, 1972b, p. 56). The reason for this improvement is a revision in the syntactic predictions (based on the previous results with other talkers) whereby boundaries are not expected (a) between pronouns and following verbals (though they are presently still predicted between non-pronoun noun phrases and following verbals) or (b) between man and the following relative pronoun who. Also, boundaries had (erroneously) not been predicted between piles and of wood, and between the adverbial conjunction till and the sun, in the earlier work.

It is expected that other refinements of the boundary predictions could be made, and should be based on a more precise theoretically-cohesive set of rules for predicting intonation contours from syntactic structure (cf. Bierwisch, 1966). A study has begun to devise such rules, incorporating some recent work of Jane Robinson (University of Michigan).

Half of the missing boundaries (predicted but not detected) were between noun phrases and following verbals, so that the planned refinement to not predict boundaries in such positions will bring the boundary detection scores to above 90%.

TABLE II.

BOUNDARY DETECTION RESULTS FOR MONOSYLLABIC SCRIPT

TALKER	NUMBER OF PREDICTED BOUNDARIES CORRECTLY DETECTED	% OF ALL PREDICTED BOUNDARIES CORRECTLY DETECTED	NUMBER OF "EXTRA" BOUNDARIES DETECTED	NUMBER OF "FALSE" BOUNDARIES DETECTED
ASH	36	86	6	7
GWH	34	80	6	3

As shown in Table II, each talker also yielded six extra boundaries, about half of which were due to "Tune 2" fundamental frequency rises at the ends of sentences (Armstrong and Ward, 1926; Lea, 1972b, p. 25). A refinement in the boundary detection procedure in sentence-final positions can readily eliminate these "Tune 2" effects (Lea, 1972b, pp. 68-69).

Seven false alarms occurred in the text by ASH, and three in that by GWH. All but one of these can be eliminated by the refinement (see discussion of the Rainbow Script) that requires that each new maximum or minimum F_o must be maintained (above the 7% threshold for F_o fall or rise) for at least 20 ms.

3.5 Boundaries Detected in the ARPA Sentences

Figure A-3 in Appendix A shows the complete boundary detection results for the thirteen ARPA Sentences, as obtained by the Univac implementation of the boundary detector. For various reasons, the overall boundary detection score (74%) is somewhat lower than for the read texts used in previous studies (79% to 90%). For one thing, some of the utterances (e.g., LM3) were quite monotone in expression, yielding insufficient F_o variations to trigger the 7% thresholds of the boundary detector. A few sentences were said with several hesitation pauses, and somewhat unusual inflections compared to the speech previously studied. As shown in Table A-I of Appendix A, the type of sentence had some effect on results, although no strong claims can be made about effects of sentence types from this small amount of data. Six of the thirteen missing boundaries were within compound nouns such as key word, Gauss elimination, and utterance eight. Despite these variations from previous results, it is encouraging that 74% of the predicted boundaries were found in these various forms of spontaneous utterances pertinent to man-machine interactions.

Extra boundaries that occurred were sometimes associated with talker's hesitations as they thought about what to say next, or with unusual stress patterns apparently associated with the spontaneity of the utterances. Eight "false" pauses occurred that were not clause or sentence boundaries, but rather were thoughtful hesitations not to be found in read speech. Some of these occurred at constituent boundaries, but not all (cf. Goldman-Eisler, 1961).

Seven false constituent boundaries were also detected, all but one of which can not be eliminated unless the minima and maxima of F_o are required to remain beyond the 7% thresholds for at least thirty milliseconds.

In Appendix A, a study is described which determined the effects of varying the threshold for "significant" F_o falls or rises. A threshold of 3% decrease or increase in F_o will allow detection of all but one predicted boundary, but will substantially increase the number of false boundaries detected, when compared to the 7% threshold value used in the present studies. These effects of threshold were very similar to those previously found (Lea, 1972b, Figure 2-5) for other texts, except that somewhat smaller F_o variations appears to be adequate for boundary marking in the spontaneous speech of the ARPA sentences. Intonational variations thus appear to be more "animated" (i.e., larger) in the reading of texts than in simulated man-machine interactions.

The Univac implementation of the constituent boundary detector allows different thresholds for F_o decreases and increases, a refinement not incorporated into Lea's earlier algorithm. The threshold studies reported in Appendix A show that better boundary detection results (that is, more predicted boundaries are actually detected while fewer false boundaries are detected) when the threshold for F_o fall is greater than that for F_o rise. This is consistent with previous studies that have shown a general trend toward falling F_o contours, with local interruptions of that falling contour marking the beginnings of new constituents.

3.6 Summary of Boundary Detection Results

Table III summarizes all the boundary detection results for the three texts, showing percents of all predicted boundaries that were detected, the numbers of extra boundaries related to minor constituent breaks, and the numbers of false boundaries.

These results encourage one to use F_o-detected boundaries in detecting significant aspects of sentence structure directly from acoustic data. This

TABLE III.

SUMMARY OF
BOUNDARY DETECTION SCORES

Text	Percent Predicted Boundaries Detected	Number of Predicted Boundaries that Were Detected	Number of Extra Detected Boundaries	Number of False Detected Boundaries
Rainbow Script: 6 Sentences 6 Talkers 252 Boundaries	79%	198	29	9
Monosyllabic Script: 5 Sentences 2 Talkers 84 Boundaries	83%	70	12	10
ARPA Sentences: 13 Sentences Mixed Talkers 50 Boundaries	74%	37	9	7+8 pauses

is true even for spontaneous utterances taken from man-machine interactions, such as the ARPA sentences. In section 5, we shall see that the successes (and occasional failures) of the boundary detector play critical roles in the process of stressed syllable location.

4. PERCEIVED STRESS PATTERNS

4.1 Experimental Methods

Experiments have also been conducted to study the stress patterns in the Rainbow Script, Monosyllabic Script, and ARPA Sentences. Actually, a three-fold experimental effort is involved in the total study of stress patterns (cf. Lea, Medress, and Skinner, 1972a). One aspect is the presentation of the scripts to individual listeners, who are asked to mark their personal judgments as to which syllables are stressed, unstressed, or reduced. A second aspect of the studies of stress is the analysis of acoustic correlates of stress, and the testing of an algorithm for stressed syllable location from acoustic data. A third aspect of the stress studies is the prediction of stress levels and vowel reductions from linguistic analyses, including syntactic analyses of the sentences in the speech texts, followed by application of appropriate stress rules and vowel reduction rules. These linguistic predictions of stress may be done with any of several available sets of rules for English stress assignment.

Only the first two aspects of these stress studies will be discussed in this report. The linguistic predictions are the subject for a future report. The algorithmic location of stressed syllables from acoustic data will be discussed in section 5. Here we consider the experiments on perceived stress.

Listeners' perceptions of stress provide a standard by which stress detections from acoustic cues can be tested. Previous studies have attempted to determine how listeners' judgments of stress vary as certain acoustic features are varied, usually in synthesized speech (cf. Lea, Medress, and Skinner, 1972a, pp. 32-40). However, few such studies have been concerned with the stress patterns throughout sentences; most work was done on isolated words such as minimal pairs of noun versus verb (permit/permit, etc.). Some attempts have been made to determine listeners' perceptions of the most stressed syllable in a sentence, or which of two specific syllables is more stressed, or whether a specific single syllable is or is not stressed. The present experiments extend studies to all syllables in the sentences.

Three listeners (WAL, MFM, and TES) each individually heard (through earphones) the Rainbow Script as recorded by the six talkers, the Monosyllabic Script as recorded by the two talkers, and the ARPA Sentences. Each listener heard clauses or sentences, or other extended portions of the text, repeated at will, by the listener's rewinding and replaying of the tape. The Rainbow Script was specifically separated into clauses separated by long pauses, to aid the rewinding and replay, while the other recordings were not. The listeners endeavored to rewind far enough to always hear an entire clause, to have a constant context within which to judge relative stress levels. Each listener could listen to the tape portions as often as necessary to mark each syllable. He was free to back up the tape at his choice, and no time limit or procedural constraints were placed on him.

The listener was instructed to mark (in whatever way he chose), for each syllable, whether he heard that syllable as stressed, unstressed, or reduced. To facilitate marking for each syllable, each script was typed on a sheet of paper with vertical slashes between syllables (except for the Monosyllabic Script, in which each word is one syllable). A mark was required for each syllable (between two slash marks). The listener received one such sheet for each talker and text. An example perception sheet is shown in Figure B1 of Appendix B.

Each listener repeated the perception test three times (with no less than three days between trials) to establish listener consistency from one time to another. Also, to establish that the actual speech heard was playing a role in stress judgments, the listeners were also asked to report their stress judgments given only the written text. This test with no speech was included to determine whether the listener's presuppositions, internal "theory" of expected stress patterns, or own way of speaking the sentences was the sole source of his decisions, or whether the acoustic data actually was supplying cues to stress patterns. These no-speech stress judgments were also obtained in three repetitions, spaced three or more days apart, to test their repeatability.

The Rainbow Script contains 127 syllables, the Monosyllabic Script 87 syllables, and the ARPA Sentences 171 syllables (71 in 6ARPA, 100 in 7 ARPA). With three repetitions with speech, three without speech, three listeners, and with the various speakers involved, this totals to about 13,000 judgments of stress levels for syllables in connected texts. In the following sections we will try to summarize these extensive results. Section 4.2 presents the majority judgments of the panel of listeners about the stress levels of all syllables in the texts. The differences between the perceptions of each listener and those of the other listeners will be explored in section 4.3. The differences from one repetition of the experiment to another will be presented in section 4.4. In section 4.5, stress perceptions from speech recordings are contrasted to stress judgments given only the written text, and implications about the English speaker-listener's rules for linguistic stress assignment are considered. Some effects of sentence type (yes-no question, WH-question, command, declarative, etc.) on stress perceptions will be discussed in section 4.6. A summary of conclusions from these stress-perception studies will be given in section 4.7.

4.2 Majority Judgments of Stress Levels

To provide a single overall decision about the stress level of each syllable in each of the texts, majority votes had to be obtained. First, for each listener, his majority vote as to the stress level of each syllable was found from comparing his three repetitions of the listening test with each text. (These judgments of the individual listener will be explored in more detail later.) Then the results for all three listeners were pooled, as shown in the plots of Figure 2 (and Figures B-2 to B-15 in Appendix B). Plotted in Figure 2, for each syllable in the Rainbow Script read by ASH, are the number of listeners whose majority vote says the syllable is stressed, minus the number of majority judgments characterizing the syllable as reduced. Unstressed judgments were assigned values of zero. Thus, if all three listeners heard a syllable as stressed (on their majority decisions from three trials), a value of +3 was plotted; if two listeners gave majority votes of reduced for a syllable, and the other listener perceived it as unstressed, a value of -2 (minus two) resulted. Occasionally (actually, very

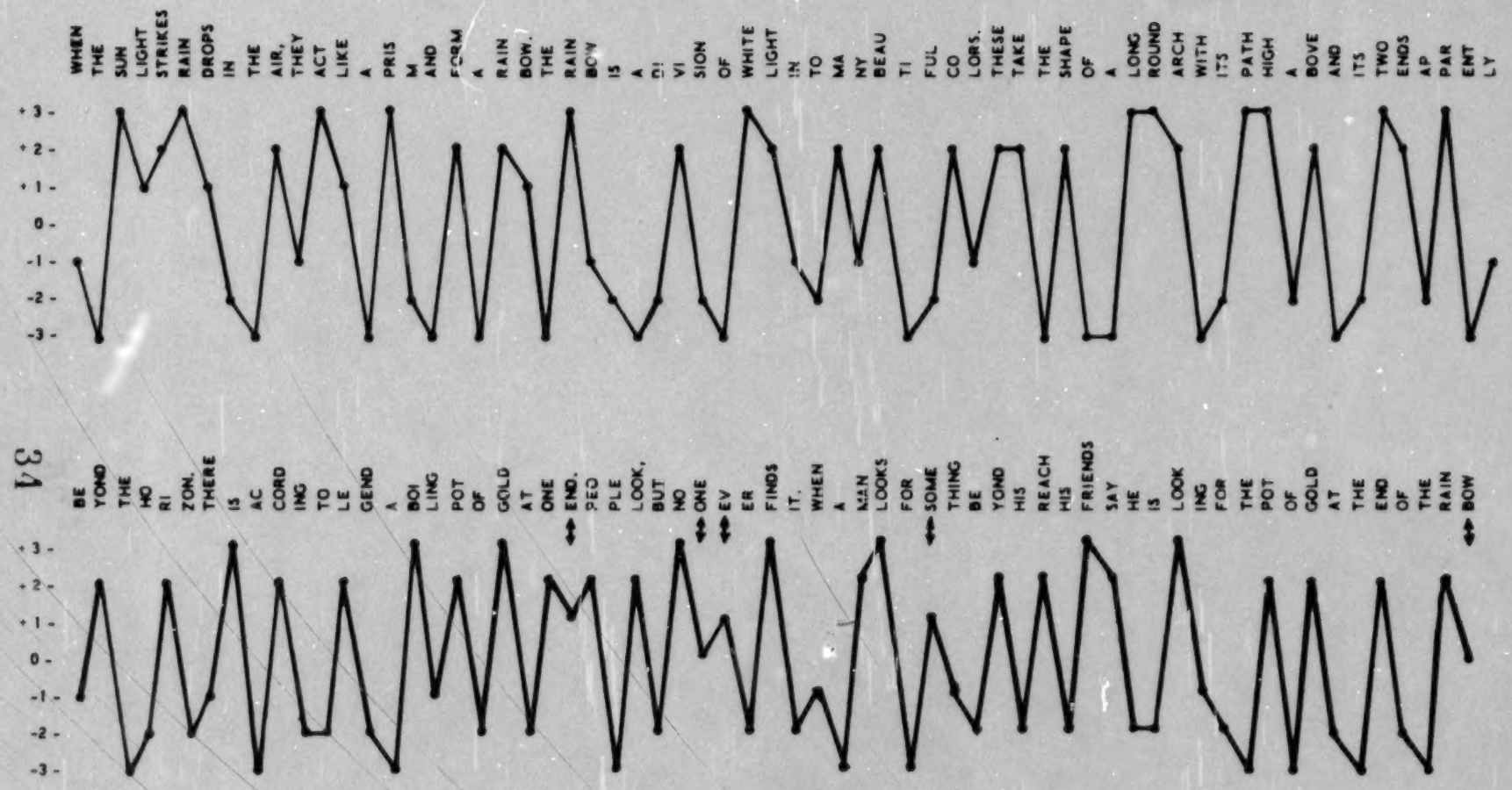

WHEN THE SUN LIGHT STRIKES RAIN DROPS IN THE AIR, THEY ACT LIKE A PRIS M AND FORM A RAIN BOW. THE RAIN BOW IS A DI VI SION OF WHITE LIGHT IN TO MA NY BEAU TI FUL CO LORS. THESE TAKE THE SHAPE OF A LONG ROUND ARCH WITH ITS PATH HIGH A BOVE AND ITS TWO ENDS AP PAR ENT LY
+3
+2
+1
0
-1
-2
-3
BE YOND THE HO RI ZON, THERE IS AC CORD ING TO LE GEND A BOI LING POT OF GOLD AT ONE END. PEO PLE LOOK, BUT NO ONE EV ER FINDS IT. WHEN A MAN LOOKS FOR SOME THING BE YOND HIS REACH HIS FRIENDS SAY HE IS LOOK ING FOR THE POT OF GOLD AT THE END OF THE RAIN BOW

Figure 2. Summary of Stress Judgments by Three Listeners, for One Talker (ASH) Reading the Rainbow Script. Plotted for each syllable is the number of listeners whose majority judgments (from three trials) declare the syllable as stressed minus the number of such majority judgments of the syllable as reduced. Unanimous judgment as stressed thus yields the top value of +3, whereas judgments as reduced pull the value down toward -3.

rarely), one listener's judgment of reduced cancelled another's judgment of the syllable as stressed. These cases of opposing judgments are marked on Figure 2 (and Figures B-2 to B-15) by double-ended arrows (↕) below the corresponding syllable of the text.

The syllables which were most definitely stressed (i.e., perceived by all listeners as stressed) thus were at the top of the scale; those definitely perceived reduced were at the bottom of the scale. From such results, one can readily see which syllables are unanimously judged as stressed, which are judged as stressed by a majority of the listeners, etc. When syllables such as long, round, path in the second sentence shown in Figure 2 are unanimously judged as stressed, one can be more confident that acoustic cues to stress are to be found. In section 5, we shall assume that all syllables which had an overall stress score of +2 or +3 are stressed, and should be found by the algorithm for stressed syllable location.

From Figures 2 and B-2 to B-15 (in Appendix B), one can observe that about 40% of all syllables were judged as stressed (stress score (SS) of +2 or +3) by the panel of listeners. About 25% were judged unstressed (SS = +1, 0, or -1), and about 35% were judged reduced (SS = -2 or -3).

Thus, if one were to analyze only the stressed syllables, as suggested in section 1, the distinctive-features analysis could be avoided in the 60% of unstressed and reduced syllables, where distinctive-features analysis is presumably more difficult and unreliable.

4.3 Effects of the Individual Listener on Stress Perceptions

It is obvious from the plots of stress scores in Figure 2 and Figures B-2 to B-15 that listeners often differ in their judgments of stress levels. Here we consider those differences in some detail.

Suppose we first consider the syllable-by-syllable differences in majority stress judgments between the listeners. (We consider here the majority decisions from three trials by one listener, compared to corresponding majority

judgments from three trials by another listener.) Every time a syllable is called stressed by one listener and unstressed by another, we have what we might call a listener-to-listener "confusion" in stress levels. Similarly, one listener's judgment as reduced and another listener's judgment as unstressed (or even stressed) represents a confusion. All of these differences in assigned stress levels can be summarized in confusion matrices, such as previously illustrated by Lea, Medress, and Skinner (1972a,b). With so many texts, talkers, and listeners, the number of confusion matrices is extremely large (but they are available for those who may be interested in studying them). The primary conclusions are, however, summarized in the plots of Figures 3 and 4.

Figure 3 shows the percentages of all stress level judgments that differ from one listener to another, plotted for each text and talker and for both conditions of SPEECH (listeners hearing the speech recordings) and NO SPEECH (individuals judging the stress levels from the written text only). There is little variation in the percentage of confusions between listeners for different texts and talkers, and speech versus no-speech conditions. However, there is a prominent effect due to the individual listener. Listeners WAL and MFM show different judgments for about 20 to 30% of all syllables. These confusions are considerably fewer than those between listeners WAL and TES (30 to 55%) and between MFM and TES (about 45 to 60%). It is apparent that listener TES gives results that are markedly different from those of the other two listeners. Listeners WAL and MFM are more alike.

Figure 4 illustrates an even more serious way in which listener TES differs from listeners WAL and MFM. Confusions (from listener to listener) between stressed and unstressed syllables are separated from those between unstressed and reduced syllables. The white bars show the percentages of unstressed-reduced confusions for each text, talker, and condition. The cross-hatched bars show corresponding percentages of confusions between stressed and unstressed levels. The extreme confusions between stressed and reduced are shown by dark bars. Listener TES actually labelled as reduced some syllables which the other listeners called stressed.

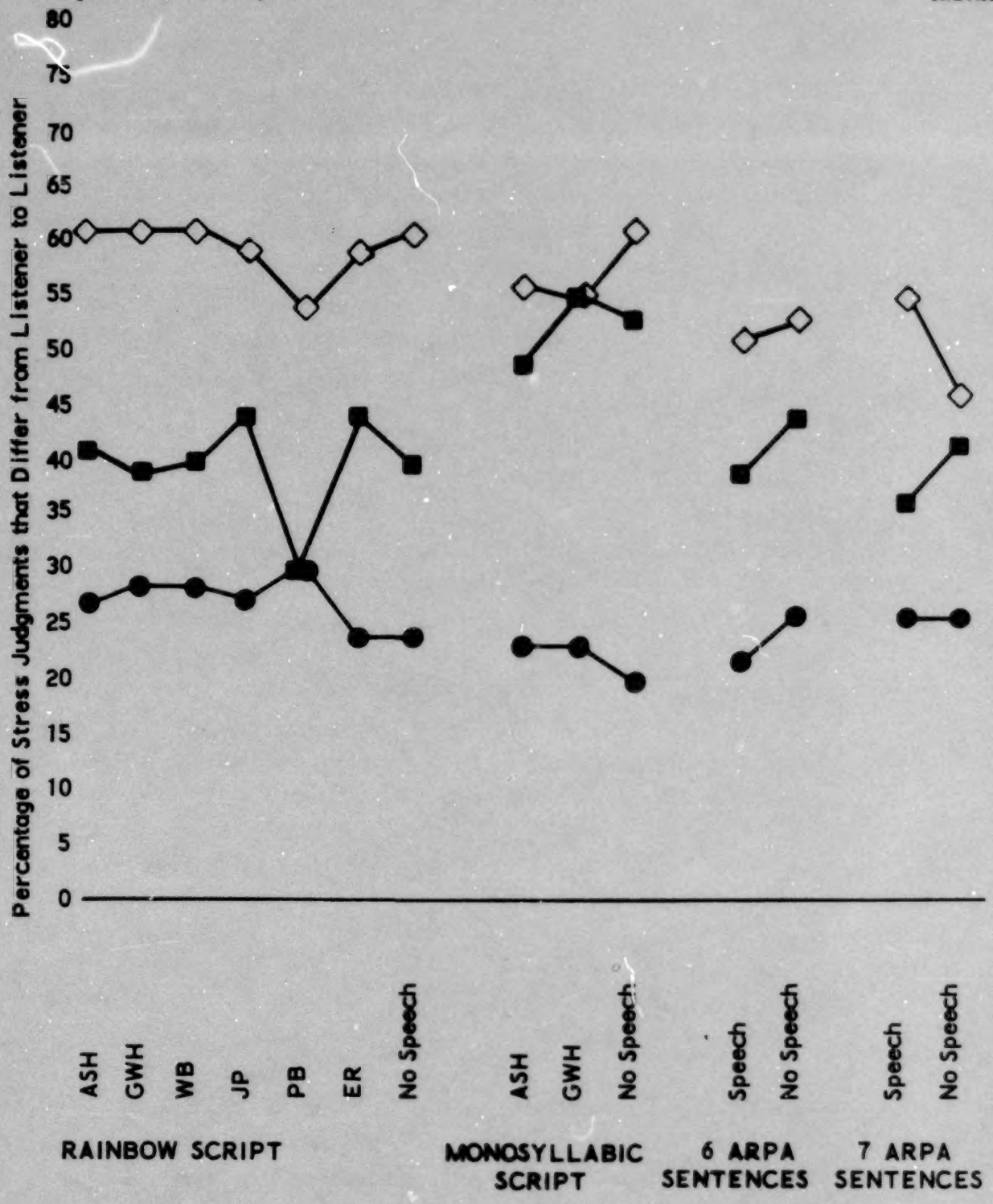


Figure 3. Percentages of Stress Judgments that Differ from One Listener To Another, for Each Speech Text, and with Each Speaker and the NO SPEECH Conditions. Plotted are percentages of confusions between listeners WAL versus MFM (-●-●-), WAL versus TES (-■-■-), and MFM versus TES (-◇-◇-).

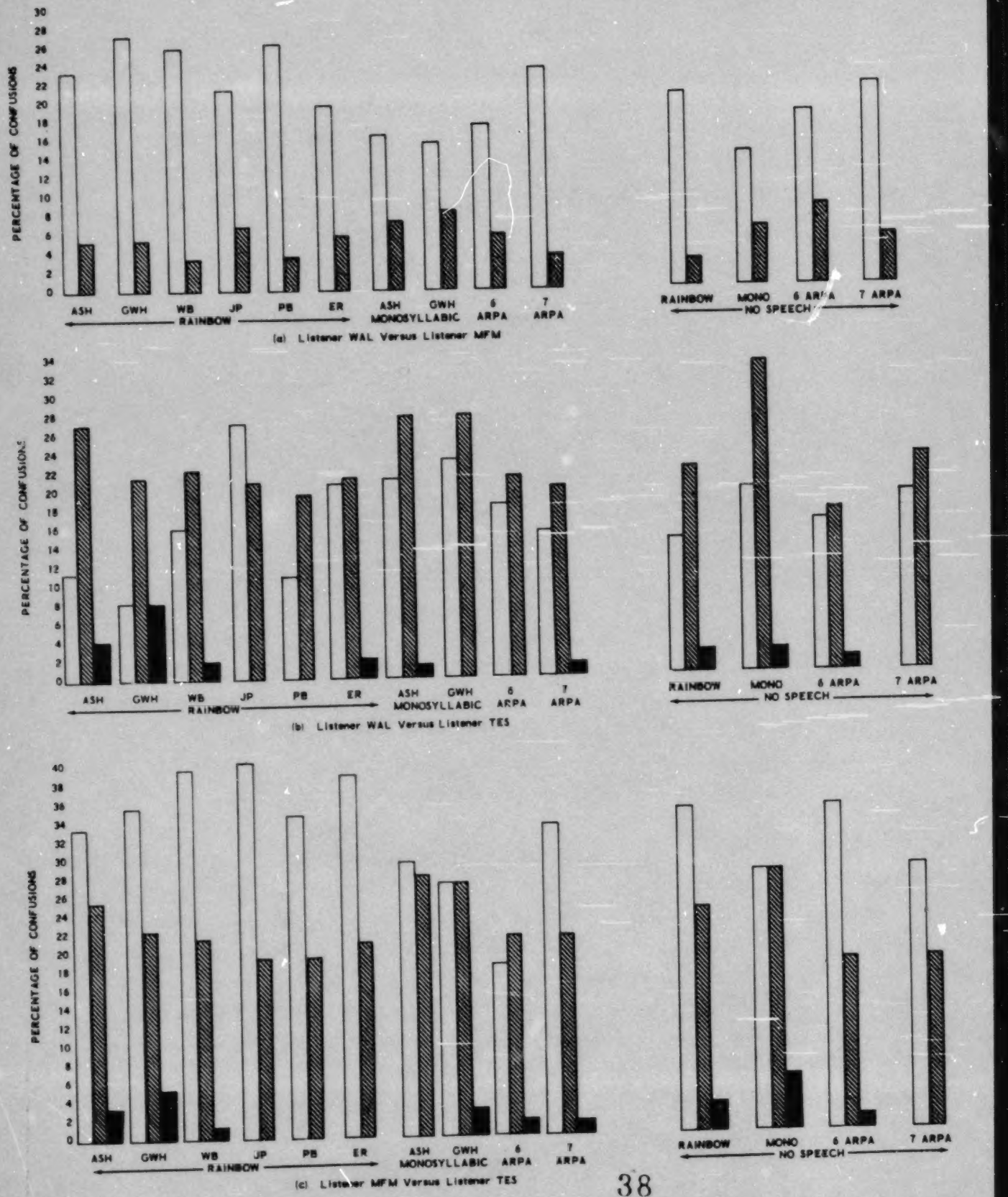


Figure 4. Percentages of Listener-to-Listener Confusions in Assigned Stress Levels for Each Text and Talker, with Unstressed-Reduced (□), Stressed-Unstressed (▨), and Stressed-Reduced (■) Confusions Separately Graphed

The important fact shown by Figure 4 is that only about 2 to 8% of all syllables were judged as stressed by listener WAL and unstressed by MFM, or vice versa, while stressed-unstressed confusions were much more frequent between listener TES and either of the other two listeners (17 to 31% for WAL vs TES, and about 18% to 28% for MFM vs TES). This is critical since listeners' judgments of stressed syllables will be used to evaluate the algorithm for stressed syllable location.

These frequent differences in assignment of stressedness to syllables, and the occasional extreme confusions between stressed and reduced syllables, suggest that one must be very careful how he pools the results for listener TES with those for listeners WAL and MFM. Our procedure for overall stress assignment by adding the stress scores for each listener yields a result which assigns stress to a syllable (for comparison with the location algorithm) whenever WAL and MFM agree that it is stressed, except in the extreme case where TES calls that same syllable reduced. (TES never called a syllable stressed when either of the other listeners didn't call it stressed.)

These differences between listener TES and other listeners were observed early in our stress perception studies (Lea, Medress, and Skinner, 1972b). Listeners WAL and MFM also were found to yield results quite similar to those of four other listeners used in previous studies at Purdue University, while TES gave quite different results. However, for consistency, the experiments were continued maintaining the same three listeners throughout.

A reasonable conclusion might be to reject listener TES. Yet, one might argue that it is conceivable that TES is actually giving the judgments closest to the "true" stress levels of syllables, and the other listeners are wrong and should be rejected. Lacking any way of deciding "true" stress levels, how does one decide the issue? After all, as has been pointed out in previous reports (Lea, Medress, and Skinner, 1972a,b), listener TES is much more demanding about the characteristics of a stressed syllable. His strategy of stress classification demanded that a syllable be very prominent before it was classified as stressed. Such syllables will presumably have the most

marked acoustic correlates of high energy, high and rising F_0, and long durations. Thus, an algorithm for stressed syllable location should be more successful in finding the fewer number of syllables that he categorizes as "stressed" than in finding all those categorized as stressed by less demanding listeners. It is then easier to get high "hit" rates in stressed syllable locations using TES's judgments. We have chosen to take the more challenging goal of finding all syllables that were judged stressed by a majority of the listeners.

In section 4.4, evidence will be given that does suggest that listener TES be rejected, not just because of his differences from other listeners, but also because listener TES is not as consistent from repetition to repetition of the experiment.

It may be useful to "screen" listeners for future experiments, to determine their consistency from repetition to repetition and their general similarity to other listeners. The stability of results shown in Figures 3 and 4 regardless of text or talker suggest that such screening might be done with a minimum amount of speech, such as one or two talkers reading one or two short texts.

4.4 Consistency of Perceptions From Time to Time

Stress perceptions were obtained from several trials by each listener, for each text and talker. to establish listener consistency from time to time. Thus, for example, listener WAL might listen once to talker ASH reading the Rainbow Script, then listen to the same tape again several (three or more) days later, then listen a third time after another few days. Periods between trials varied from as few as three days to as long as six or seven months. Results were reasonably consistent regardless of the period between trials, provided that the period was one week or more. For some trials separated by only a few days, the listeners reported that they could remember some of their previous assignments. Future studies should require a minimum of one week between trials.

Figure 5 illustrates the percentages of all judgments that differ from one trial to another. This is compiled for each text and talker, and for the NO SPEECH conditions, using the following procedure. For a given recording, the perceptions on trial A are compared to those for trial B. For each syllable that they differ (such as syllable air being judged stressed on one trial and unstressed on the other), one confusion would be shown off the main diagonal of a confusion matrix. The number of syllables whose two trial judgments differ (yielding off-diagonal instances in the trial A versus trial B confusion matrix), divided by the number of syllables in the text, gives the percentages of syllables confused from trial A to trial B. This is repeated for trial B versus trial C, and for trial A versus trial C. The averages of these three percentages of (off-diagonal) confusions is the value plotted for each text and talker in Figure 5. Results are shown separately for each listener's confusions from trial to trial. There is thus a very large amount of confusion data summarized in Figure 5.

The results show that listeners are fairly consistent from trial to trial, regardless of text or talker. That is, less than 24% of all judgments vary from trial to trial. For the Rainbow Script and the 7ARPA Sentences, results are quite similar from listener to listener and from talker to talker, or even from talker to NO SPEECH conditions. However, listener TES yielded considerably more trial-to-trial confusions than listeners WAL and MFM for the Monosyllabic Script, where his more frequent stressed-unstressed confusions were undoubtedly affected by the many stressed syllables occurring in texts of monosyllabic words. Trial-to-trial confusions were particularly numerous in the 6ARPA sentences, especially for NO SPEECH conditions. We shall see later that this was in part due to the questions and unusual pauses and F_o variations involved in these spontaneous utterances.

Figure 6 presents a breakdown of repetition-to-repetition confusions into those between stressed and unstressed, unstressed and reduced, and stressed and reduced, for each listener. As in Figure 4, where listener-to-listener confusions were plotted, it is apparent that listeners WAL and MFM showed few

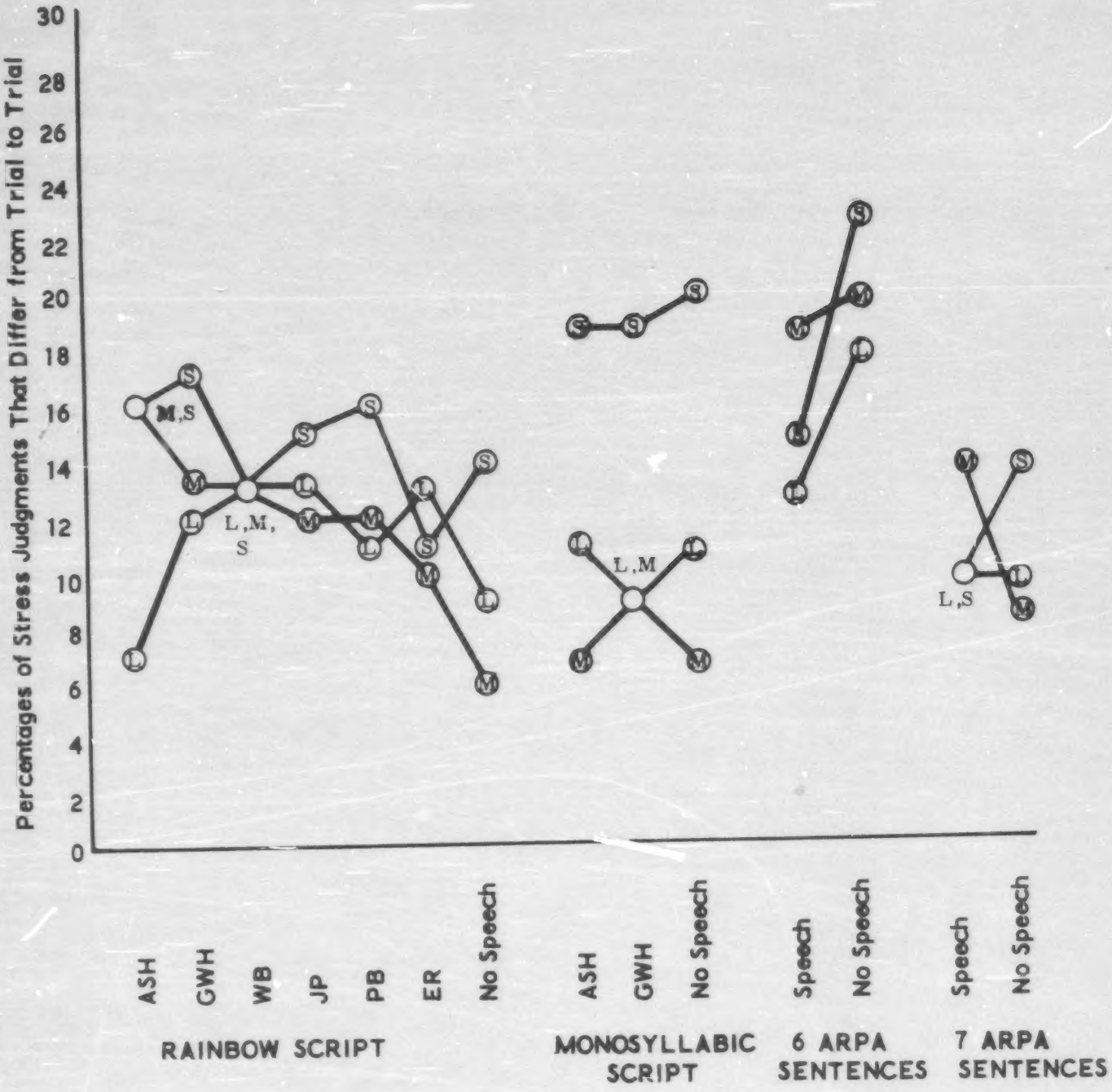
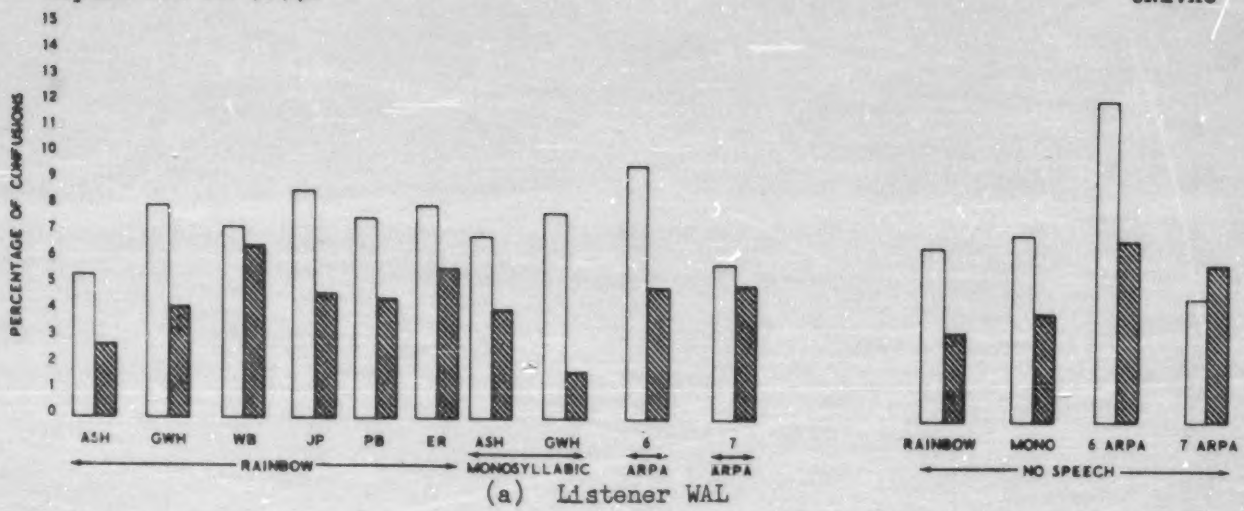
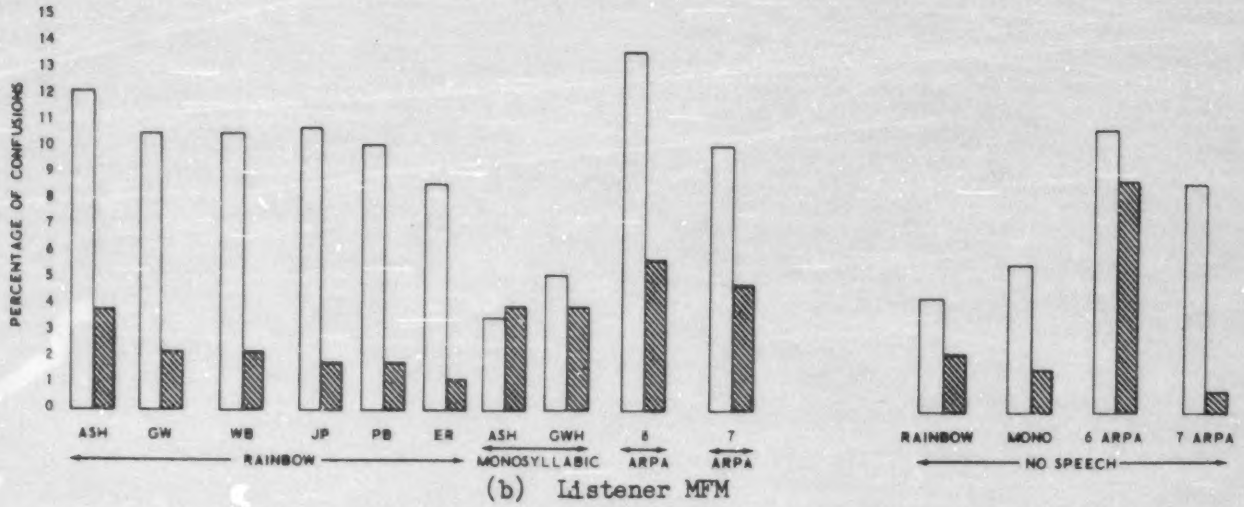


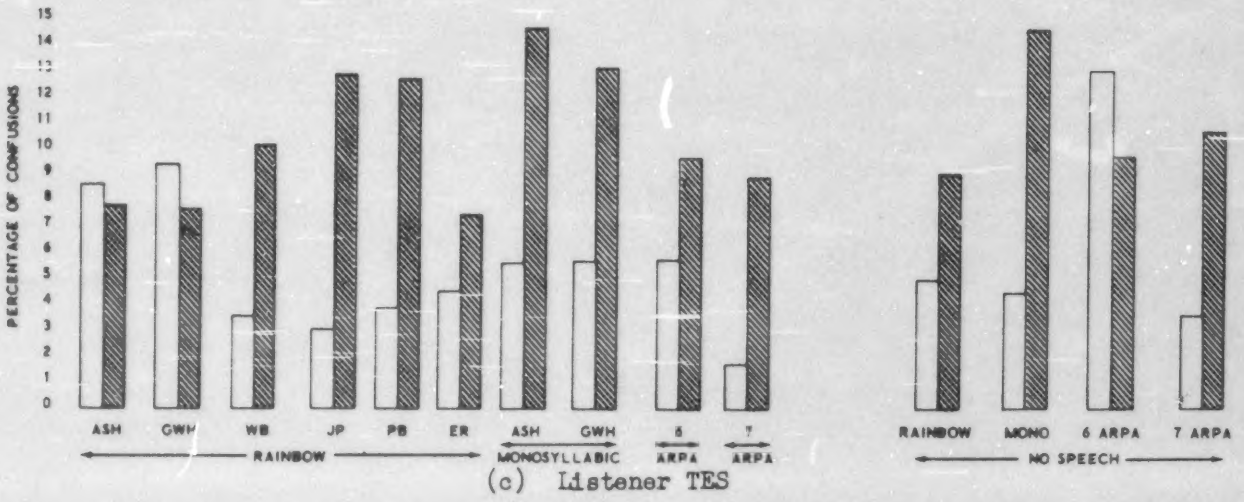
Figure 5. Percentages of Stress Judgments that Differ from One Trial to Another, for Each Speech Text, and with Each Speaker and the NO SPEECH Conditions. Plotted are percentages of confusions from Trial to Trial for listener WAL (Ⓛ), listener MFM (Ⓜ), and listener TES (Ⓢ).


PERCENTAGE OF CONFUSIONS
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
ASH GWH WB JP PB ER ASH GWH 6 7
RAINBOW
MONOSYLLABIC
6 ARPA
7 ARPA
RAINBOW MONO 6 ARPA 7 ARPA
NO SPEECH

(a) Listener WAL


PERCENTAGE OF CONFUSIONS
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
ASH GW WB JP PB ER ASH GWH 6 7
RAINBOW
MONOSYLLABIC
6 ARPA
7 ARPA
RAINBOW MONO 6 ARPA 7 ARPA
NO SPEECH

(b) Listener MFM


PERCENTAGE OF CONFUSIONS
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
ASH GWH WB JP PB ER ASH GWH 6 7
RAINBOW
MONOSYLLABIC
6 ARPA
7 ARPA
RAINBOW MONO 6 ARPA 7 ARPA
NO SPEECH

(c) Listener TES

Figure 6. Percentages of Repetition-To-Repetition Confusions in Assigned Stress Levels by Each Listener, for Each Text and Talker, with Unstressed-Reduced ☐ and Stressed-Unstressed ▨ Confusions Separately Graphed.

(1 to 9%) stressed-unstressed (and no stressed-reduced) confusions, while listener TES gave many more confusions (7% to 14%) from trial-to-trial. In fact, listener TES produced more stressed-unstressed confusions than unstressed-reduced ones. Since the primary intent of the stress perception studies is to provide stress standards by which a stressed-syllable locator may be judged, such confusions about stressed syllables are crucial. The lack of repeatability in stress judgments, when coupled with the other unusual characteristics of TES judgments, would seem to be unacceptable in future studies of stressed syllables.

We shall see in section 5 that the stressed syllable location algorithm locates about 85% of all syllables perceived as stressed by the majority of listeners. It, thus, misses about 15% of the stressed syllables, and it labels about 15% of the syllables as stressed even though they were not perceived as stressed by two or more listeners. When the perception "standard" whereby the algorithm is judged varies from time to time by the same order of magnitude as the differences between the perceptions and the acoustically-derived decisions, it can hardly be called a "standard" any more. We desire that the past results with the standard accurately predict the next results when applying that standard again to the measurement of the same data. We, thus, must reject TES data for providing an evaluation of stressed syllable location to any closer than 10 or 15% or so.

Even with the perceptions of listeners WAL and MFM, we must realize that the confusions of about 5% or so from time to time suggest we can not judge the effectiveness of stressed syllable location to any precision greater than about 5%. If a stressed syllable algorithm locates 95% of all syllables perceived as stressed by majority votes of two or more listeners, it is doing no worse than one repetition of the perceptions would do for predicting the perceptions from a second repetition of the experiment. We thus have no motivation to attain 98% "correct" location of stressed syllables versus 95%, etc., as long as we use the present form of listener perceptions as the standard.

One might speculate that a new procedure for obtaining listener judgments of stress levels, such as allowing a scale of 1 to 10, or an assignment of any arbitrary number to each syllable, might conceivably yield improved (more stable) perception results. However, it is doubtful that increasing the number of levels into which stress is categorized will actually improve the stability of results. A confusion of level 6 and 7 (on a 10-level scale) from repetition to repetition is still a confusion even though it may be said to be a finer-grained, or smaller, confusion than a stressed-unstressed confusion. One might try to define metrics for measuring the size of such confusions, and try to suggest that the overall confusion is decreased in some sense. However, it is important to realize that an experiment so defined does not define an interval measurement scale, in the measurement-theoretic sense (Stevens, 1951; 1969; Lea, 1971), and no such metrics would be justified in terms of the abstract structure of the perceptual scale. The present experiments define an ordinal measurement-theoric scale, which distinguishes three nominal classes (stressed, unstressed, reduced) with an ordering (stressed is "greater" than unstressed, unstressed is "greater" than reduced), but no defined intervals (we have not required or demonstrated that the "distance" or difference from stressed to unstressed is equal to that from unstressed to reduced, etc.).

Since confusions do occur from repetition to repetition of the stress perception experiment, majority votes from three or more trials would seem to be suitable for obtaining somwhat more stable results. The majority votes from three trials are expected to be more like those majorities from three more subsequent trials than the single trial-to-trial judgments would be.

4.5 Comparing Stress Judgments With Speech to Those Without Speech

The general consistency with which most listeners can assign stress levels to syllables in connected speech (Li, Hughes, and Snow, 1972; Lea, Medress, and Skinner, 1972a) suggests that there is indeed some psychological reality to the concept of stress. The fact that listeners assign

approximately the same stress patterns to the speech of different talkers reading the same text suggests that either (a) the talkers are all consistently conveying something that we might call the normative, unmarked pattern of linguistic stress for that structure and content of the text, or (b) the listeners are assigning stress levels not so much on the basis of this stable input acoustic data, but rather on the basis of their own internalized theories of stress or their projection of how they would have said the same text.

Some evidence is already available to discount the idea that the acoustic data plays absolutely no role in stress perceptions. Previous research on acoustic correlates of stress have shown that listeners do change their stress judgments as acoustic parameters are varied under various controlled conditions (cf. e.g., Lieberman, 1960; 1967; Lehiste, 1970). The data in the present experiments (see Figures B-2 to B-15 in Appendix B) show some differences from talker to talker, for the same text, which are consistently shown in the assigned stress levels of all listeners. The listener is indeed making his stress judgments based at least in part on the acoustic data, and not simply on the basis of expected patterns. It would appear that talkers are generally assigning equivalent stress patterns to the texts they speak, presumably following an unmarked "linguistic" stress pattern determined by the lexical content and structure of the sentence, but that the individual talker will occasionally deviate from a strict pattern, perhaps assigning added emphasis to certain words, or reducing certain syllables one time whereas he (or someone else) may not do quite the same thing the next time he spoke the same text.

If the talkers did in fact approximate to, but not always exactly attain to, a standard linguistic stress pattern, and if the listeners used their own internal notions of stress and the acoustic data to assign stress levels, but were not perfectly consistent in setting the boundaries between the categories of stressed, unstressed, and reduced, we might expect all the general consistencies and minor inconsistencies that have been found

in perceptions of various listeners from repetition to repetition, talker to talker, and text to text. We might expect somewhat more agreement about stress in read speech than in spontaneous utterances if the listener's a priori way of assigning stress agreed more with the unmarked pattern expected in reading texts than with the possible special emphasis, reductions, pausing and restarting and other variations that are introduced by spontaneous speech. If the listener were making stress judgments entirely on the basis of acoustic data, and had no added difficulty in making acoustic distinctions for spontaneous speech, we would expect his judgments for spontaneous and read speech to be equally consistent.

Further stress studies are needed to answer many questions about how listeners perceive stress, how their own internal models interact with the acoustic data, whether there is a consistent normative or unmarked stress pattern used by both talkers and listeners, how spontaneous speech might differ from read speech in spoken and perceived stress patterns, etc.

Included in the present studies were experiments on stress judgments given only the written text, which were to be compared with the same person's stress perceptions using the speech recordings. These NO SPEECH judgments have been included in the summaries of Figures 3 to 6 with no previous attempt to contrast them with the results with speech. Here we specifically explore the differences and similarities that result.

The listener-to-listener confusions in stress levels, as shown in Figure 3 (and in Figure 4), show no marked differences between numbers of confused perceptions with speech recordings to numbers of confused judgments without speech. We might have expected that if the listener's own stress theory (or his own way of assigning stress to the text if he were to read it) were playing an active, dominant role in stress assignment from the written text alone, and if his theory played much less of a role in listening to the speech recordings, and if the internal theories of the listeners differed much, then the listener-to-listener confusions without speech should be substantially more than those with the equalizing

effect of the acoustic data. But, in fact, there is no significant difference between the percentages of listener-to-listener confusions with versus without speech. Thus, either the listeners are each assigning substantially the same stress patterns whether the speech is present or not (and thus some internal theory is playing a dominant role under both conditions) or else they all vary in similar manners in how they change no-speech judgments to perceptions with speech.

Suppose one could show that stress judgments exhibit many more confusions from repetition to repetition when only the written text is given, when compared to the perception confusions from repetition to repetition with speech. Then he could argue that the present stress perception experiments using speech recordings are more useful than just having native English subjects predict stress patterns from the written text. He could also argue that this is evidence that the acoustic data were critical in obtaining reliable stress assignments. Surprisingly, this did not turn out to be true in the present experiments! Figures 5 and 6 show that, with the possible exception of the results for the 6ARPA Sentences, the number of repetition-to-repetition confusions without speech is not significantly larger than the number of confusions with the speech.

A related issue is whether the stress judgments without speech agree substantially with the perceptions with speech. That is, can one accurately predict the listener's perceptions with speech from his judgments without speech (or vice versa)? While judgments without speech may be consistent from time to time, and while numbers of listener-to-listener confusions may be comparable with or without speech, the syllable-by-syllable judgments without speech may or may not correspond with those with speech. Figure 7 illustrates the results of comparing the majority decisions (for three trials) of each listener with speech to his majority decisions without speech. Plotted are the percentages of all syllables in the texts that are assigned different stress levels with speech from those assigned without speech, for each listener.

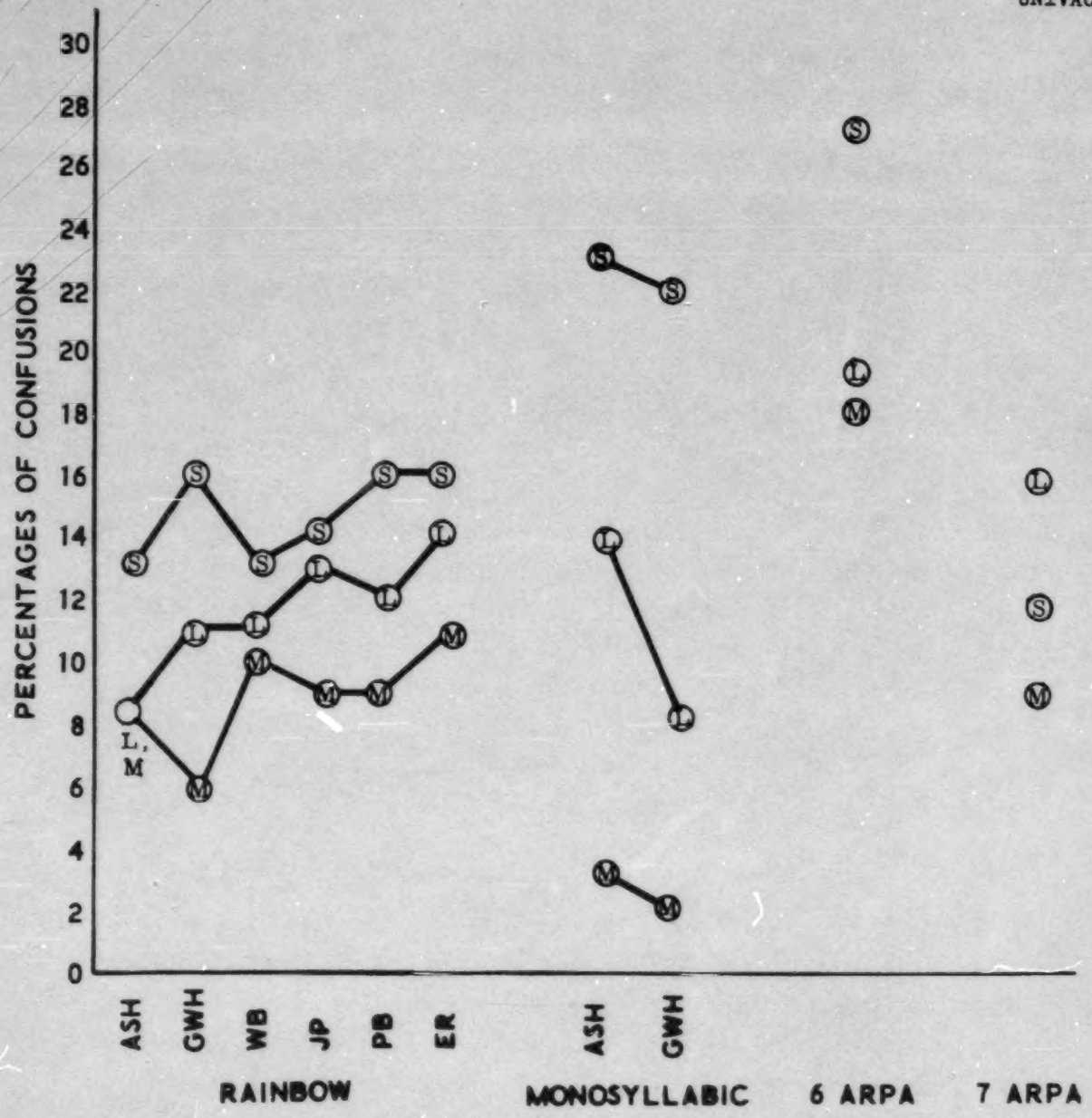


Figure 7. Percentages of Confusions in Assigned Stress Levels for NO-SPEECH versus SPEECH Conditions, for Each Text and Talker. Plotted are percentages of confusions for listener WAL (Ⓛ), listener MFM (Ⓜ), and listener TES (Ⓢ).

It is evident from comparing the results of Figures 7 and 5 that, for the Rainbow Script, comparable percentages of confusions occur for repetition-to-repetition comparisons (Figure 5) and speech-to-no speech comparisons (Figure 7). That is, the NO-SPEECH stress judgments do as good a job of predicting stress perceptions with speech as one repetition with speech would do in predicting the results of another repetition with speech, for the Rainbow Script. For listener MFM, the majority NO SPEECH judgments for most texts are more like the majority judgments with speech than one repetition with speech is like the next repetition with speech. On the other hand, listeners WAL and TES usually show more confusions between SPEECH and NO SPEECH than between repetitions with speech, particularly for the Monosyllabic Script and the ARPA Sentences. (The probable reason these listeners did not show more SPEECH vs NO SPEECH confusions for the Rainbow Script is that for that text, the NO-SPEECH judgments were done after the listeners had done three tests with the speech, and discussed the results, so their NO-SPEECH judgments could have been biased by previous experience with the speech. For the Monosyllabic Script, the NO-SPEECH judgments were obtained before any test with the speech. For the ARPA Sentences, some NO-SPEECH tests were performed before, and some after, the tests with speech. Data analyses for those texts were done after all experiments had been performed.)

The vast difference in SPEECH vs NO SPEECH confusions for the Monosyllabic Script might suggest that listeners vary in their relative success of assigning lexical versus structure-dictated aspects of stress. Listener MFM shows very little (2% or 3%) confusions for the Monosyllabic Script, perhaps indicating that he can assign sentence stress very consistently. His higher rates of confusion (6% to 18%) with other texts (notably, the 6ARPA Sentences) suggest that he has more difficulty when lexical stress factors of polysyllabic words also are involved. Listener TES, on the other hand, is quite inconsistent in assigning stress to the Monosyllabic Script, perhaps suggesting more difficulty with sentence structure aspects of stress assignment. An equally revealing and perhaps more plausible explanation is that the Monosyllabic Script has a higher percentage of stressed syllables than the other texts. We have already seen that MFM has considerably fewer stressed-unstressed confusions than

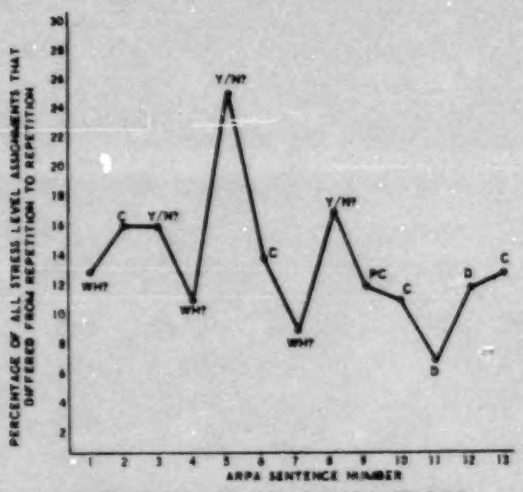
unstressed-reduced, while TES confuses considerably more stressed and unstressed syllables.

The relatively high confusion rates shown in Figure 7 for the 6ARPA Sentences (and in Figure 5 for NO SPEECH confusions from repetition-to-repetition) suggest that we cannot rely on stress judgments using only the written text to give the best predictions of perceived stress levels for spontaneous utterances suitable for man-machine interactions. Thus, while stress judgments without speech recordings may do a surprisingly good job of predicting perceived stress patterns for normal speech read from texts, they are not the best form of stress judgments for spontaneous speech. Stress location algorithms to be used in speech understanding systems should be judged by stress perceptions obtained from speech recordings, not from judgments about orthographic transcriptions.

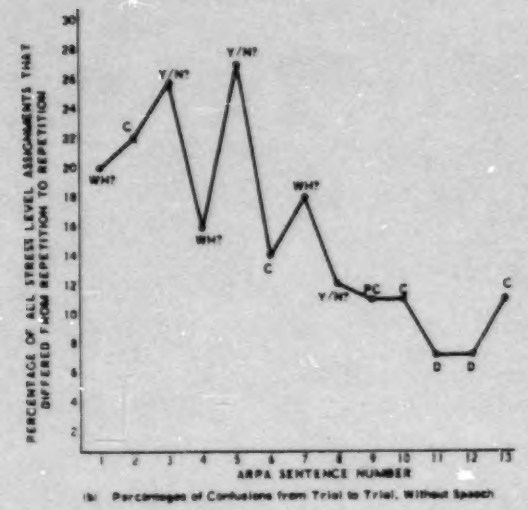
4.6 Effects of Sentence Type on Stress Judgments

In compiling the confusions for the various texts, it was evident not only that confusions were more common in the 6ARPA Sentences, but that questions seemed to exhibit more confusions than declaratives or commands. In Figure 8, the thirteen ARPA sentences are separately listed, along with symbols that indicate the basic category to which that sentence structure belongs (yes-no question, Y/N?; question with interrogative (WH) word, WH?; command, C; polite command, PC; and declarative, D). Plotted for each sentence is the percent of all syllable stress level comparisons that differed from repetition-to-repetition, for the three trials with speech (Figure 8a) or without speech (Figure 8b), or the percent of all syllables that differed between the majority vote with speech and the majority vote without speech (Figure 8c). Results were pooled for all listeners, by first finding the plot for each individual listener, then averaging the values for all three listeners for each sentence.

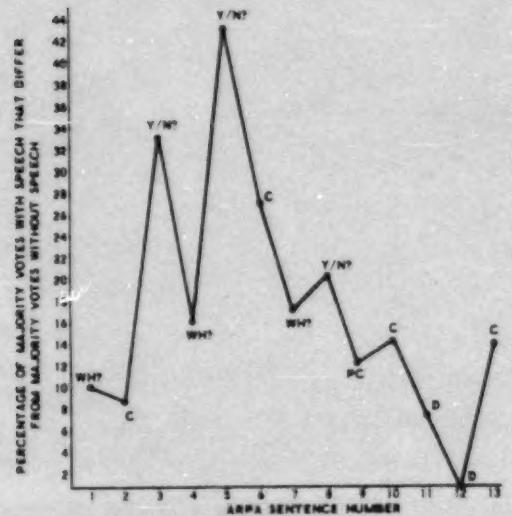
Several points may be made from these results. First, percentages of confusions tend to be higher for the 6ARPA Sentences than for the 7ARPA Sentences in Figure 8b and c, presumably because the NO SPEECH judgments



(a) Percentages of Confusions from Trial to Trial, With Speech



(b) Percentages of Confusions from Trial to Trial, Without Speech



(c) Percentages of Cases Where the Majority Vote from Three Trials With Speech Differed from the Majority Vote from Three Trials Without Speech

Figure 6. Effects of Individual Sentence Type on the Percentages of Stress Level Confusions for the ARPA Sentences. Plots shown are the averages of the separate plots for each of the three listeners. Sentences are identified as Yes/No Questions (Y/N?), WH Questions (WH?), Commands (C), Polite Commands (PC), and Declaratives (D).

for the 6ARPA Sentences were obtained before the listeners heard the speech, while, for the 7ARPA Sentences, some NO SPEECH judgments were made after the perceptions with the speech had been obtained. The WITH SPEECH perceptions seemed to have been remembered, to help stabilize stress decisions in later trials.

Of more importance here are the high confusion percentages that occur in Figure 8a, 8b, and 8c for Yes-No questions 3, 5, and 8, and for WH question 7 in 8c. In general, questions (especially yes-no questions) seem to exhibit more confusions than declaratives and most commands.

Another way in which stress perceptions are significantly influenced by the sentence type is in terms of how much the different listeners varied from each other in their consistency of stress assignment for each sentence. Listeners differed by 40% in the percentage of confusions which they exhibited from NO SPEECH to SPEECH for questions 7 and 8, 22% for yes-no question 3, and 20% for yes-no question 1, but less than an average of 13% for all of the other ARPA sentences. Similarly, in repetition-to-repetition confusions without speech, the greatest variations in rate of confusions occurred for yes-no questions 3 and 8, and WH questions 1 and 7 (as much as 30% compared to an average of 11% for the other ARPA Sentences).

From these preliminary results, it appears that stress assignment is more difficult in questions than in other sentence structures. Further, more controlled tests with various sentence types would be needed to confirm these apparent trends obtained from only 13 sentences. These tests will be undertaken using the extensive set of sentences presently being designed for isolating various factors affecting prosodic patterns (cf. Lea, Medress, and Skinner, 1972a, pp. 56-7).

4.7 General Conclusions About Stress Perceptions

The above extensive analyses of stress assignments by three listeners have yielded the following general conclusions:

1. Different listeners assign different stress levels to the same syllables, presumably based on how they individually define the boundaries between categories of stressed, unstressed and reduced syllables. Their confusions are not seriously increased or decreased in going from individual talker to talker, or from text to text (except when questions are introduced; see point 8 below).

2. Listeners WAL and MFM, who have been shown by previous experiments to yield stress perceptions very much like those of other listeners, differed in as much as 25 to 30% of their majority decisions about stress levels (compiled from three trials). However, only about 5% of all syllables were confused between the categories stressed and unstressed. Thus, judgments of which syllables were stressed agreed very well between listener WAL and listener MFM.

3. Listener TES differed from the other two listeners on about half of his stress decisions. About 20 to 25% of all syllables were labelled stressed by other listeners, but unstressed by TES. He actually even labelled as reduced some syllables labelled stressed by the other listeners. Also, listener TES labelled substantial percentages (as much as 15%) of all syllables as stressed on one trial and unstressed on another. Future studies should incorporate a procedure for rejecting such listeners who provide inconsistent judgments about stressed syllables.

4. From repetition to repetition of the perception tests, listeners WAL and MFM individually showed quite stable judgments as to which syllables were stressed. An average of 5% of all syllables were confused between stressed on one trial and unstressed on another trial. They thus provide a reasonably stable "standard" as to which syllables are stressed, for comparison with algorithm results.

5. Majority votes obtained from 3 or more trials should be used to partially obliterate the 5% deviations in assignment of stressed syllables from trial to trial. No stressed syllable location algorithm need find more than 95% of all syllables perceived as stressed, since it can hardly be more "accurate" than one perception trial is in predicting the perceptions to be attained on another trial.

6. Since listeners agreed in many of the differences they assigned to the stress patterns of different talkers reading the same text, the acoustic data appears to play at least some role in stress perceptions. However, since listener-to-listener confusions and most repetition-to-repetition confusions were not significantly increased when only the written text was used, it appears that the listeners also make use of a reasonably stable internal theory for stress assignment.

7. When listeners had not done the perception tests with speech before they did the stress assignments from the written text alone (as with the Monosyllabic Script and the ARPA Sentences), their majority judgments without speech differed more from their majority perceptions with speech than the repetition-to-repetition with speech (or without speech) had differed. Thus, while stress judgments without speech are as consistent from listener-to-listener and from repetition-to-repetition as are perceptions with speech, the judgments made without speech are significantly different from those made with speech. In particular, perceived stress patterns for spontaneous utterances are not reliably obtained from judgments based only on the written text.

8. Questions (especially yes-no questions) appear to yield more confusions in stress levels (from repetition-to-repetition) than other sentence structures (declaratives or commands), and show greater variability from listener-to-listener.

In summary, the stress perceptions obtained from the trials with speech, by using majority decisions for each listener, and pooling results for the listeners by the sum (-3 to +3) plots as shown in Figure 1, provide a "standard" of stress assignment which is stable to within about 5%. This thus permits comparisons (to within 5%) between perceived stressed syllables and stressed syllables located by algorithm from the acoustic data.

5. STRESSED SYLLABLE LOCATION FROM ACOUSTIC DATA

5.1 Correlates of Stress in F_o Contours

Stress is an abstract quantity usually considered to be associated with a speaker's total physical effort in speech production or with a listener's perception of "prominent" syllables. Having obtained extensive data on listeners' perceptions of stressed, unstressed, and reduced syllables (section 4), we shall now consider how the perceptions relate to acoustic data.

Extensive work has been done on acoustic correlates of stress (cf. reviews by Lehiste, 1970, and Medress and Skinner, 1972), and on physiological correlates of stressed syllable production (cf. review by Lieberman, 1967). Many studies have taken advantage of the ability to separately control acoustic features of synthesized speech, to test how acoustic variations correlate with listeners' perceptions of stress (Fry, 1955, 1958; Bolinger, 1958; Morton and Jassem, 1965; Mattingly, 1966; etc.). Most experimental studies have been concerned with stress in isolated words, short phrases, or short isolated sentences.

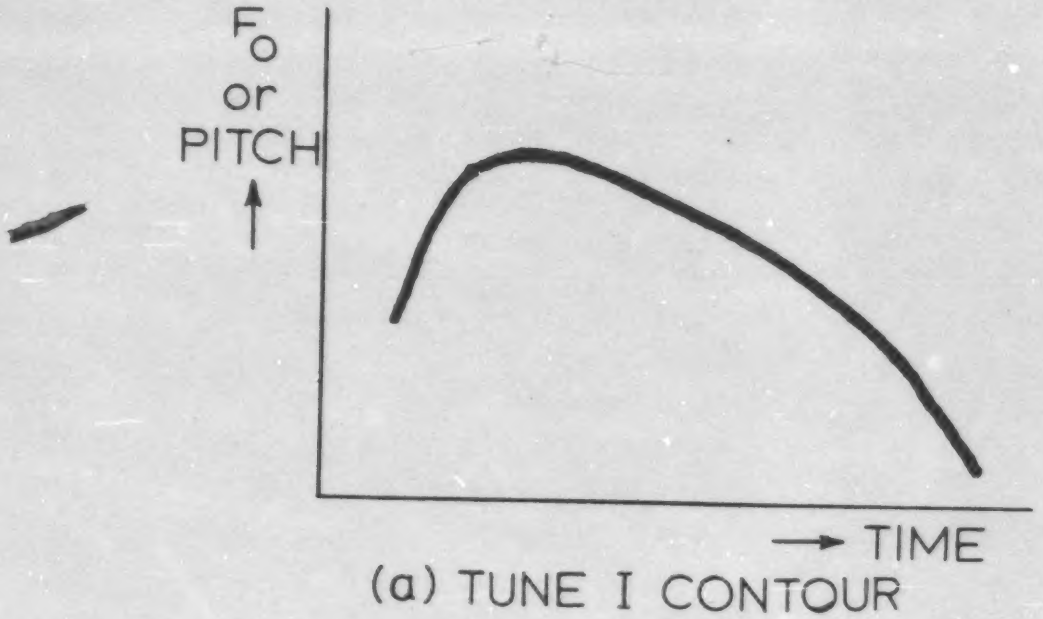
For reasons detailed previously (Lea, Medress, and Skinner, 1972a), the stress perception studies reported in section 4, and the studies of acoustic data reported in this section, are concerned with stress patterns in semantically-connected texts and computer instructions or queries, spoken by several different native English speakers. Acoustic correlates of stress that will be incorporated into the stressed syllable location algorithm are (1) fundamental frequency (F_o) variations (particularly local increases in fundamental frequency) and (2) the energy integral within the syllable (incorporating both amplitude and duration measures into one measurement).

While many studies have shown that higher F_o is associated with stressed syllables (Bolinger, 1958; Fry, 1958; Lieberman, 1960; Morton and Jassem, 1965; Lehiste, 1970; Lea, 1972b), others have shown that even better correspondence is to be found between local increases (or, occasionally, decreases) in F_o and

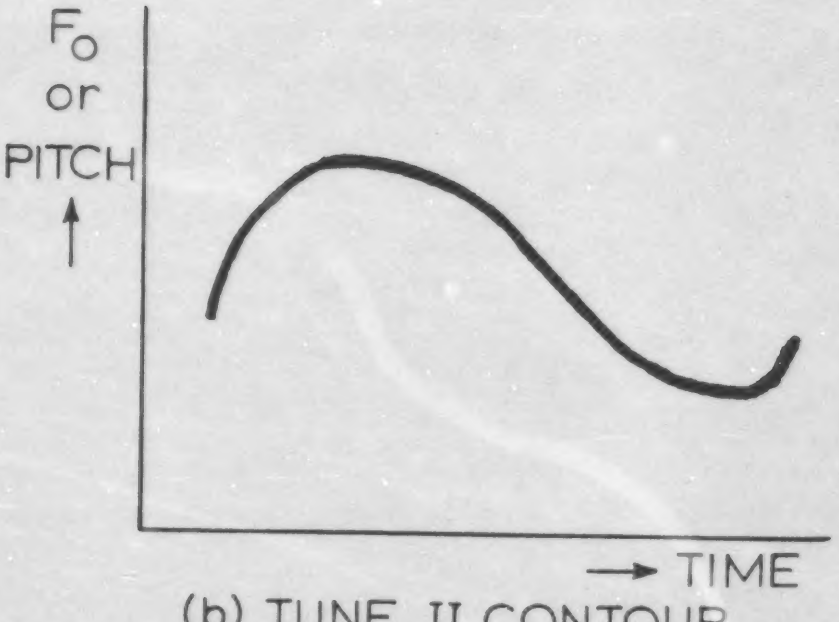
stress then is provided by the absolute peak (or mean) values of F_o within stressed vowels or syllables (Bolinger, 1958; Medress and Skinner, 1971; Morton and Jassem, 1965). Some studies suggest that it is the presence of such F_o changes that marks stress, not the specific magnitude of the change (Fry, 1958; Morton and Jassem, 1965).

Effects of phonetic sequences may interfere with these stress effects on F_o contours. Vowels articulated with higher tongue position have F_o values that are about 10 to 15% higher than those with low tongue position (House and Fairbanks, 1952; Lehiste, 1970; Lea, 1972b, 1973a). This is one reason why absolute values of F_o may fail to mark stress; an unstressed /i/ may have a higher peak or mean F_o than a somewhat more stressed /a/. Peak or mean F_o in a vowel is higher when the preceding consonant is unvoiced than if it is voiced or if no consonant precedes the vowel (House and Fairbanks, 1952; Lea, 1972b). More important with respect to the F_o changes associated with stress are the sudden F_o changes that occur around consonants (cf. Lea, 1972b, Chapters 4 and 5). Fundamental frequency suddenly drops about 10% within the closure period of voiced obstruents, suddenly rises again at opening of the closure, and continues to rise (about 15% or more) during the 100 ms after the following vowel onset. For unvoiced consonants, F_o will cease (sometimes after the 10% dip at closure, since voicing frequently ceases after closure) and then, when voicing resumes, F_o will start quite high and rapidly fall. These dips and sudden cusps in F_o contours must somehow be distinguished from stress-dictated F_o changes.

Another influence on F_o contours must be considered in establishing acoustic correlates of stress. Intonation studies (Armstrong and Ward, 1926; Lieberman, 1967; Lea, 1972b) have shown that, in connected texts and spoken sentences, F_o will usually reach a maximum near the first stressed syllable (the so-called "HEAD") of each breath group or clause, and will fall gradually until the last stressed syllable, after which may occur either the rapid fall of an utterance-final "Tune I" contour or the rise in F_o at the end of "Tune II" contours (which mark "incompletion"). Figure 9 illustrates the general shapes of these basic intonation contours. Obviously, the last stressed



(a) TUNE I CONTOUR



(b) TUNE II CONTOUR

Figure 9. Tune I and Tune II Intonation Contours

syllable of Tune I contours will not consistently exhibit the F_o rises generally assumed to accompany stressed syllables. Also, unstressed syllables in the terminal rise of a Tune II contour will be accompanied by F_o rises that do not mark stress. On the other hand, these studies suggest that the peak F_o of the contour will be associated with a stressed syllable.

The assumption of the constituent boundary detector is that sentences consisting of several major grammatical constituents will be broken into several Tune I- or Tune II-like subcontours, riding on the general tune for a sentence or clause. Thus, as illustrated in Figure 10, F_o contours in sentences with several major constituents will have major F_o changes associated with the constituent structure. We might call these rapidly rising and gradually falling F_o contours as "archetype constituent contours". They resemble Lieberman's (1967) unmarked and marked breath groups, and Pike's (1945) primary contours plus precontours, and other contours associated with "sense groups" in the literature.

We shall build a general hypothesis about F_o correlates of stress based on archetype contours within constituents. It appears the rising F_o near the beginning of a constituent is attributable to the first stressed syllable in the constituent (Lea, Medress, and Skinner, 1972b). An algorithm for stressed syllable location should thus search in the region of the peak F_o in the constituent, and the rising F_o region proceding the peak. In fact, it appears that the F_o rise that marks the beginning of the "constituent" found by the boundary detector is associated with this following stressed syllable. In a sense, then, the constituent boundary detector may be said to be detecting some stressed syllables (but not locating them). If each constituent had exactly one lexical word with a major-stressed syllable within it (as has been suggested for deep structures; cf. Chomsky, 1965; Emonds, 1970), we might expect the present method of constituent detections to be closely associated with the presence of stressed syllables.

In fact, however, surface structure constituents (both as predicted syntactically and as found by the boundary detector) sometimes contain more

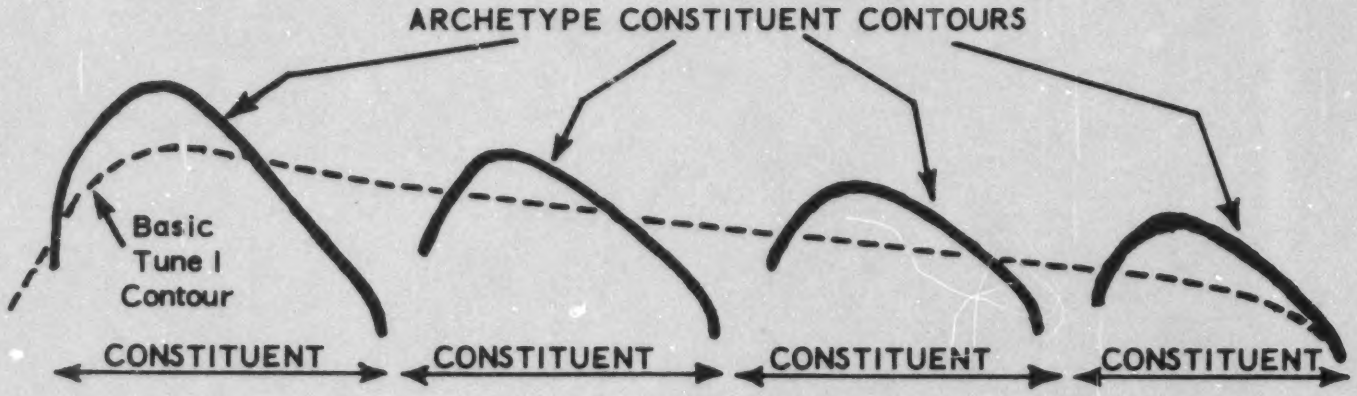


Figure 10. Each Major Constituent of a Sentence Is Assumed To Exhibit a Rapidly-Rising, Gradually-Falling "Archetype Constituent Contour", Riding on the Overall Tune I Contour of the Sentence. Tune II-like subcontours (with brief terminal rises in F_0) may replace the unmarked archetype constituent contours when incompletion is explicitly marked in the constituent.

than one stressed syllable (as in the constituent into many beautiful colors in the Rainbow Script). Based on previous studies showing higher F_o and rising F_o to be associated with stressed syllables, we might expect that the extra stressed syllables in the constituent will be accompanied by local increases in F_o, above the general archetype pattern. Since these additional stressed syllables are assumed to follow the first stressed syllable associated with the peak F_o, any increases in F_o associated with them will be manifested by bumps (temporary increases in F_o) above the archetype falling F_o contours, as shown in Figure 11.

This general strategy regarding F_o correlates of stress will not detect all stressed syllables in all of speech. When special emphasis, specific "marked" semantic attitudes (such as unbelief, distrust, etc.), or other non-normative non-neutral expression forms are intended by the speaker, he may show sudden decreases in F_o on stressed syllables (cf. Pike, 1945; Bolinger, 1958; Lea, Medress, and Skinner, 1972a, pp. 35-6). Also, some constituent structures do not always show highest F_o on the first stressed syllable in a constituent, but rather on some later stressed syllables. This will introduce cases where a stressed syllable is not located by an algorithm based on the archetype contours.

5.2 Energy-Integral Cues to Stress

Early studies of acoustic correlates of stress showed that vowel durations were longer, and vowel intensities were higher in stressed syllables (Fry, 1955; 1958; Lieberman, 1960). Indeed, the earliest works (Saussure, 1915; Jones, 1932) equated high intensity and stressedness. However, later studies showed that F_o was usually the best of the three individual correlates (Fry, 1958; Lieberman, 1960; Bolinger, 1958). Then, Lieberman (1960) showed that the energy values integrated over the vowel or the total syllabic duration gave the best cue to stressed syllables. Medress and Skinner (1971) found that the energy integral (over the vowel) was the strongest cue to stress, most successfully determining the stressed vowel in multisyllabic words, both in isolation and in short sentences.

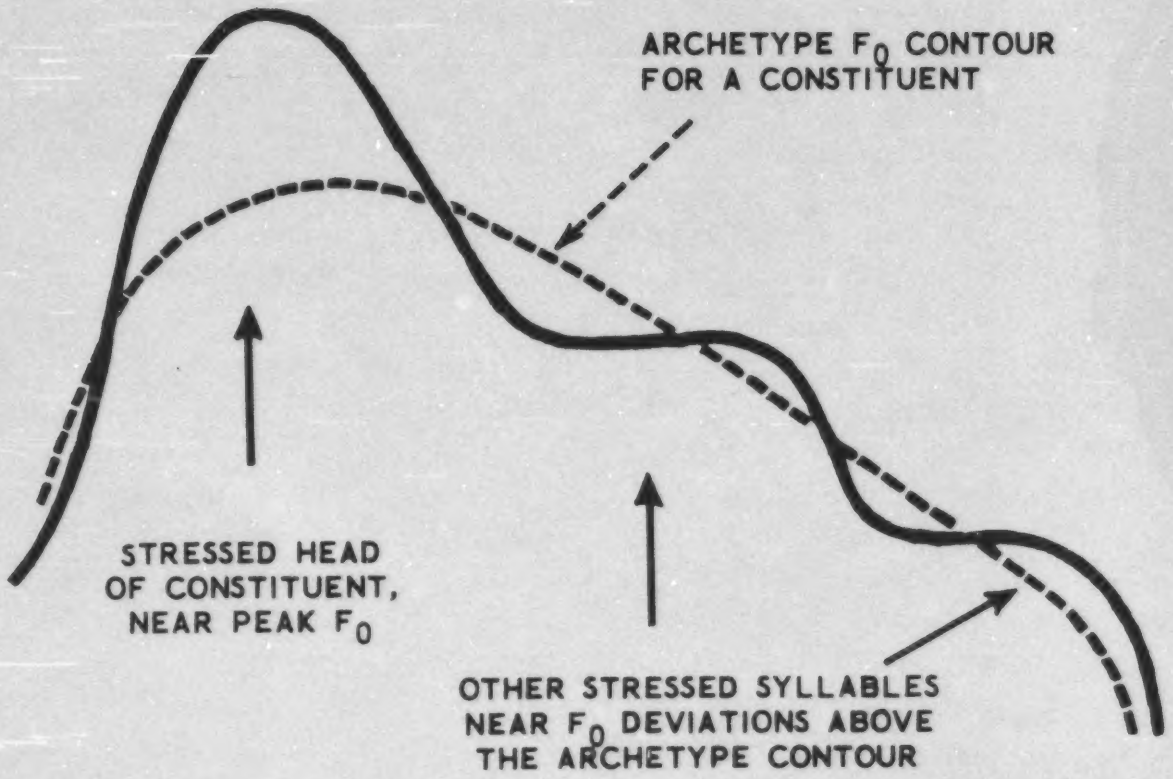


Figure 11. Increases in F_0, Above the Archetype Contour for a Constituent, Are Assumed to be Associated with Stressed Syllables.

The energy integral, which incorporates both durations of integration and intensitities at each point within that period, is affected by phonetic content of the words, and by the positions within intonation contours of total structures. Vowels articulated with low tongue positions, such as /æ , a/ are more intense (by as much as 6 db) and longer (by as much as 25%) than those with high tongue positions, such as /i, u/ (House and Fairbanks, 1952; Lehiste, 1970). Tense vowels are also longer than lax vowels (Delattre, 1962). Vowels are longer when followed by voiced consonants than when followed by unvoiced consonants. Vowels in unvoiced consonantal environments tend to be less intense (House and Fairbanks, 1952). The manner of articulation of following consonants can also affect the durations of vowels. Finally, word- or phrase-initial vowels tend to be more intense than word-final, phrase-final or utterance-final ones, while phrase-final syllables tend to be longer in duration than medial or initial syllables (Lehiste, 1970; Mattingly, 1966). The phrase-final (or so-called "prepausal") lengthening of syllables appears to be different for stressed and unstressed syllables (Oller, 1971).

Morton and Jassem (1965) showed that about 6 db or more is needed between the intensity levels of syllables to successfully mark stress. Intensity variations of 3 db or less are insignificant perceptually. Syllabic nuclei (vowels and prevocalic or postvocalic non-vowel consonants) are at least 6 db more intense than intersyllabic consonants. Thus, syllabic segmentation of speech would presumably involve 6 db variations in intensity.

Based on these various studies of duration and intensity, and their relationships to stress, a general strategy of stressed syllable location from energy integrals can be outlined. Within the constituents detected by the boundary detector, and near the positions of peak F_o and local increases in F_o above the archetype contour, a search should be made for periods of high intensity, yielding large energy integrals, bounded by dips in energy presumed to mark syllabic boundaries. These dips should

be on the order of 6 db. Given several high energy regions in the vicinity of an F_o increase, one should select one with highest energy integral and non-falling or rising F_o.

It is conceivable that a number of detailed refinements to this general strategy could maximize the accuracy of stressed syllable location. Among such refinements could be adjustments to account for intrinsic F_o, intensity, and duration of various vowels, to account for effects of surrounding consonants, and to account for positions within total structures and intonation contours.

5.3 An Algorithm for Stressed Syllable Location

As shown on the example computer listing in Figure 12, the constituent boundary detection program provides markers ("SYNTB") for the positions of all detected syntactic boundaries, plus markers ("MAXFO") at the (first) position (time ITMAX) of maximum F_o in each constituent. These are used as starting data for the stressed syllable location algorithm. The algorithm for stressed syllable location, which is detailed in Figure C-1 of Appendix C, proceeds by first locating the HEAD stressed syllable in the constituent, then finding any other stressed syllables between the HEAD and the end of the constituent. Presently, no details are included to normalize for vowel identity, phonetic context, or position within the total intonation contour (such as utterance-final tonalization positions, etc.; cf. Lea, Medress, and Skinner, 1972a, p. 37). In this section we shall sketch some of the main points and detailed decisions involved in the stressed syllable location algorithm. A flow chart is given in Figure C-1 of Appendix C.

5.3.1 Finding the First Stressed Syllable in a Constituent

To find the HEAD of a constituent, the algorithm begins with the position ITMAX of maximum F_o in the constituent. If contiguous points after ITMAX maintain that same maximum F_o (forming a plateau), the center point of such constant-F_o points is called the "Time of Peak" (TOP). (See Figure 12, where the three segments following MAXFO maintain the same F_o value.) If, however,

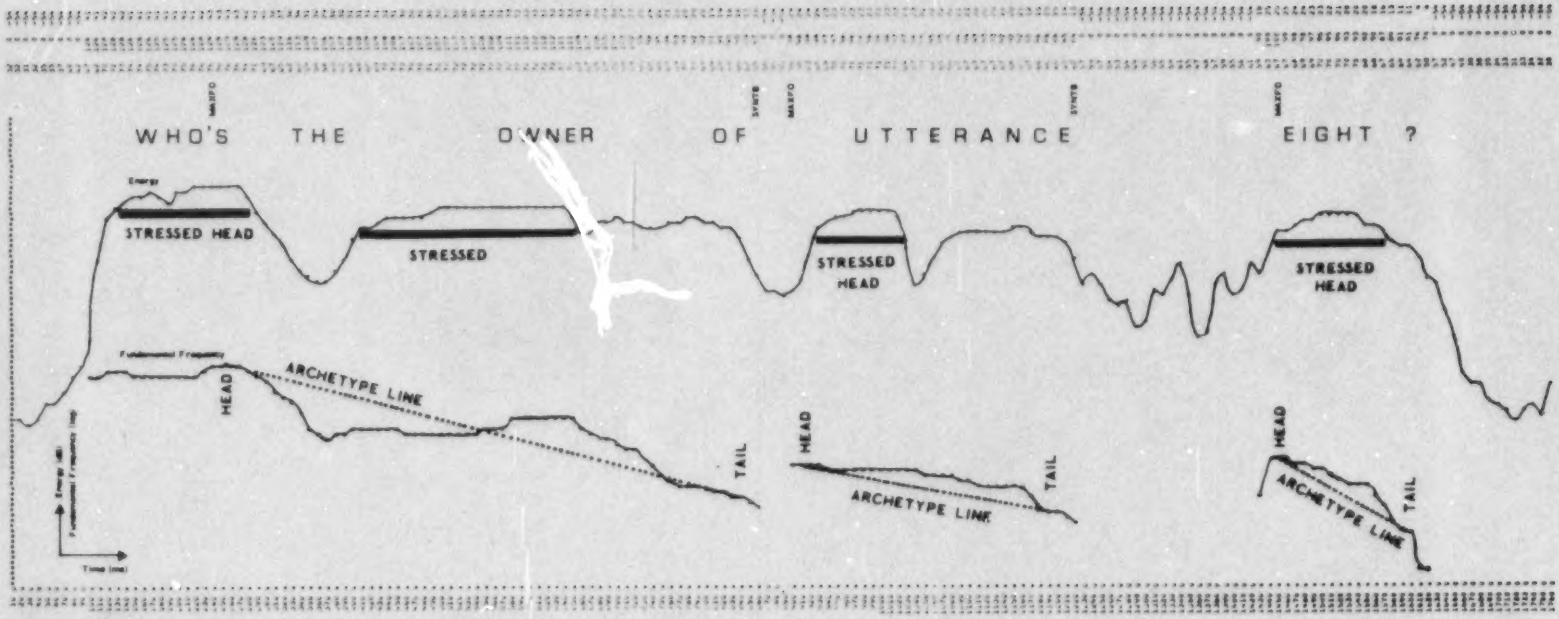


Figure 12. Computer Printout of the Fundamental Frequency and Broadband Speech Energy Functions for Each 10 ms of the Question "Who's the owner of utterance eight?" Above the plots of F_0 and energy are tabulated the values of energy in dB, F_0 in Hertz, and F_0 in eighth tones (with zero eighth tones, the bottom of the F_0 plot, set at 70 Hertz). A SYNTB marker is shown at the time of each detected syntactic boundary, and MAXFO is shown at the first occurrence of maximum F_0 within each constituent. The algorithm for stressed syllable location locates a STRESSED HEAD near the MAXFO point, then assigns an archetype line from the HEAD to the TAIL of the F_0 contour of the constituent, and locates other STRESSED portions of the speech.

F_o falls after ITMAX, and the utterance was unvoiced for two or more time segments immediately before ITMAX, then a check is made on F_o values just before the unvoicing began. If F_o before the unvoicing was within a threshold percentage (presently THMAX = 20%) of F_o at ITMAX, and if F_o had been non-falling or rising in the five time segments before unvoicing, then set TOP equal to the time of the last voiced segment just before unvoicing.

The Time of Peak (TOP) gives a reasonable starting point from which to search for the first stressed syllable. The test for previous unvoicing is to allow for the fact that F_o may be higher immediately after voicing onset after an unvoiced consonant than it is in the previous syllable, even though the previous syllable may be the more stressed of the two syllables. This refinement is also needed to provide a more reasonable starting point for the archetype falling contour to be assigned following the HEAD of the constituent. Proper setting of TOP can significantly affect the slope of the archetype following contour.

The next step is to search for the likely location of the stressed syllable near TOP which will form the HEAD of the constituent. Within some length of time BACKT (presently four hundred milliseconds) before TOP and some threshold time FORWT (now three hundred milliseconds) after TOP, a search is made for all dips in energy of a threshold amount EDIP (now set at 5 db variations for the broadband energy function defined by Lea, Medress, and Skinner, 1972a, p. 23). (The most efficient means for finding these and other energy dips used in stressed syllable location is to precede the stressed location program by a little program that finds and marks all peaks and dips in the energy contour, just as the boundary detector provides for F_o.)

If only two dips occur within the time interval defined by BACKT and FORWT, then the high energy portion bracketed by those dips will form the time temporarily assumed to be associated with the HEAD of the constituent. (With the present values of BACKT and FORWT this is highly unlikely, and the 700 ms will need to be divided into two or more syllables by other procedures

described below.) If more than two dips occur within the bracketed time of BACKT and FORWT (as is the case in the example of Figure 12), tests must be made for which high energy portion between dips is to be called the stressed HEAD.

First the energy integral ENERGY is defined for each portion between dips in the bracketed time region (portions before the first dip but after the beginning of BACKT, and after the last dip but before the end of FORWT are neglected in this comparison of energy integrals). (This energy integral specification might be most efficiently determined by the preliminary program that finds energy peaks and dips.) The energy integrals are presently found by simply summing the energy values of all time segments between the dips. Where appropriate, the relative sizes of these energy integrals may be used to select the portions which are the stressed syllables.

The present algorithm assumes a preeminence of F_o as a stress cue, so, before considering energy integrals, an F_o test is made. Of the several high energy portions within the bracketed time, find all those which exhibit an overall rise in F_o during the time that the energy does not dip below its maximum by more than 3 db. That is, F_o at the first point where energy is within 3 db of maximum (such as time segment labelled 150 in Figure 12) must be less than F_o at the last point (such as time segment 280 in Figure 12) before energy drops 3 db below the maximum. If more than two such portions have rising F_o, choose the first one unless the first one is only five or less time segments in length or unless the energy integral of (any of) the second or later one(s) is (are) markedly (presently 40% or more) greater than that for the first one. If no portions show rising F_o, then choose the highest in energy integral.

If the high energy HEAD so selected is very long (with 300 ms or more between its preceding energy dip and its following dip), then a test will be made for more than one stressed syllable within it. Sometimes two or more syllables without intervocalic obstruents will show no significant (5 db) energy dips, and would appear to form a single "stressed syllable". If

there is a small dip of at least 2 db lasting for two or more time segments, breaking the apparent HEAD into two high energy portions each of at least 100 ms in duration, and if F_o in the later portion is above the archetype F_o contour to be defined below, this second portion will be labelled as another stressed syllable, distinct from the HEAD.

5.3.2 Finding Other Stressed Syllables in a Constituent

Having found a stressed syllable corresponding to the HEAD for each constituent, the stressed syllable location algorithm next searches for other stressed syllables within each constituent. First, the TTAIL (time of the TAIL) of the F_o contour is defined as the center of the last plateau or bottom of the last small (2% or greater) valley of F_o within the constituent (such as time segment 850 in Figure 12), not including the plateau or valley bottom that the boundary position is set within. Next, a linear archetype plot on the eighth-tone (logarithmic) F_o scale is drawn from the eighth-tone value at the TOP to the eighth-tone value at the TAIL.

Then a search is made for all instances, after the energy dip marking the end of the HEAD and before the TTAIL, where the eighth-tone value of F_o for five or more consecutive segments is greater than that defined by the archetype line. (When the HEAD is longer than 300 ms so that two or more stressed syllables might be included in the HEAD, the test for increases in F_o above the archetype begins at 100 msec before the end of the HEAD, or at the small 2 db energy dip defining a possible syllable boundary, whichever is first.) If F_o in eighth-tones is above the archetype line for the minimum duration (presently set at five time segments) and if F_o is rising during that time, or flat, then the high energy portion (more than 60 ms in duration), bounded by 5 db dips which is associated with this F_o rise is called another stressed syllable in the constituent. To determine which high energy portion is associated with this non-falling F_o stretch above the archetype line (that is, to establish the bounds of this other stressed syllable), a search for nearby high energy portions is made. If no energy dips occur in the time that F_o is non-falling, then the stressed syllable extends to the immediately preceding and following 5 db dips (such as at

time segments 410 and 660 in Figure 12). If dips do occur during the non-falling portion of the F_o contour above the archetype line, the same tactics for selection apply as with HEADS; namely, the first stretch with rising F_o and high energy is chosen unless it is too short (less than 60 ms) or lower in energy integral by 40% or more than a following high energy portion whose F_o is still above the archetype.

One case is also allowed where a falling F_o which is still above the archetype line can be declared a stressed syllable. If, for six or more time segments, F_o is above the archetype line but falling, a search for 5 db energy dips in that area is undertaken. Between two dips, determine the total portion that is within 3 db of the maximum intensity. If F_o does not fall more than an average of two eighth-tones per five time segments within this high energy portion, then that portion is also declared a stressed syllable. This allows stressed syllables where F_o had been falling rapidly, but was locally increased above the archetype to be a much more gradual fall. Thus, the increase in F_o, above what might have been, really marks the presence of a stressed syllable, even if the F_o is not rising absolutely.

5.4 Comparison of Algorithmic Locations With Perceived Stress Patterns

The stressed syllable location algorithm has not been implemented as a computer program. However, it has been followed strictly in a hand analysis of acoustic cues to stress patterns for the speech texts listed in section 2. The results of such algorithmic locations of stressed syllables were compared with the perceptions of stress. The complete sets of algorithmic results are shown in Figures C2 to C11 in Appendix C. Those figures show the texts as spoken by the various talkers, with a box around each syllable that was perceived as stressed by two or more listeners (that is, that had a stress score of +2 or +3; see section 4.2). Also shown in the figures are lines underscoring all those portions of the texts that were found within the high energy portions declared as "stressed syllables" by the algorithm.

Thus, for example, Figure C3 shows that the syllables sun-, strikes, rain-, air, act, pris-, form and rain- were perceived as stressed in the

first sentence of the Rainbow Script as read by talker GWH. However, the algorithm found when, sunlight, strikes, rain-, in the air, act, pris-, and rain as included in the high energy portions declared to be "stressed syllables". Thus, it gave a "false" detection of when as stressed, missed the stressed syllable form, and included within some "stressed syllables" portions which were unstressed. Extended voiced sequences, and especially sonorant sequences, may have no significant energy dips, so that sequences such as in the air, -orizon, boiling, when a man looks, the end, etc. may be included in "stressed syllables". As long as a stressed syllable is included within each such stretch, we may consider that no false alarm has occurred, and that that stressed syllable has been correctly located. However, if two stressed syllables were included within the single stretch, we would consider one correct location (and one miss).

Stretches which the algorithm declares stressed but which did not include any syllable with a stress score of +2 or +3 are considered "false" stress detections (e.g., when in Figure C-3). One major source of such false alarms is a false boundary detection (e.g., as in the middle of the word contain in ARPA Sentence 3 shown in Figure C-10). When false boundaries are assigned, they demand that a stressed HEAD be found in each of the surrounding constituents (so that, e.g., con- must be a stressed HEAD since it is a constituent). Some located portions also occur where listeners WAL and MFM perceived a syllable as stressed, but since listener TES perceived it as reduced, it was assigned a stress score of +1. With a more consistent set of listeners, these may be perceived as stressed and the location would be correct.

The stress scores marked on the false locations and missing locations in Figures C-2 to C-11 show that many false alarms were on syllables with stress score +1 (perceived as stressed by at least one listener), while most misses were on syllables of score +2, where not all listeners agreed the syllable was stressed.

Table IV summarizes the stressed syllable location results from the hand analysis with the algorithm. Shown for each text and talker are the numbers of syllables perceived as stressed, the numbers of those found by the algorithm, and the consequent percentages of all stressed syllables that were correctly detected. Also shown are the numbers of false locations in each run, and the percentages of all locations by the algorithm that were false (that is, did not include syllables perceived as stressed).

While scores varied some from text to text and talker to talker, the overall scores of 78% to 98% (average, 85%) correct location of stressed syllables are very encouraging. The Monosyllabic Script, with its prominent stresses on monosyllabic words, yielded quite high scores. The spontaneous ARPA sentences, which were more monotone and which gave some difficulties to the boundary detection algorithm, showed the lowest stressed syllable location scores.

The false alarm rates were fairly high, ranging from 7% up to 28%. Some of the false alarms will be eliminated by improvements in the boundary detector. Some other "false" locations are not necessarily bad, since one or two listeners did perceive those syllables as stressed. A few of the false alarms may be eliminated by not demanding stressed HEADS in short constituents (such as those less than 200 ms in duration). Further studies are needed to reduce false alarm rates and simultaneously maintain or improve the scores for correct locations. Ultimately, the design of a better algorithm for stressed syllable location must be based on a strategic decision as to whether it is better to have some false alarms and correspondingly increase the success in correct location or to have little or no false alarms but at the sacrifice of lower scores in correct location. This will substantially depend upon the specific use of stressed syllable information in other aspects of the speech understanding system. For guiding distinctive features estimation procedures, all that might come from having a few false locations is that distinctive features analysis may occasionally be applied (perhaps wastefully or with some difficulty) in the somewhat-less-reliably-encoded unstressed syllables.

TABLE IV.

STRESSED SYLLABLE LOCATION SCORES

Text	Number of Stressed Syllables Perceived by Two or More Listeners (P)	Number of Such Syllables Correctly Detected (D)	Percent Stressed Syllables Correctly Detected (D/P x 100%)	Number of 'False' Locations (F)	Percent of All Locations That Were 'False' (F/D+ F) x 100%
RAINBOW					
ASH	51	43	84%	3	7%
GWH	45	44	98%	7	14%
WB	47	38	81%	15	28%
JP	48	42	88%	15	26%
PB	50	39	78%	6	13%
ER	49	43	88%	3	7%
MONOSYLLABIC					
ASH	41	37	90%	8	18%
GWH	41	39	95%	14	26%
13 ARPA SENTENCES	70	56	80%	14	20%

It would be of interest to compare the substantial success in stressed syllable location which was attained with the present algorithm with results that might be attained with other algorithms, such as simpler ones that merely look for all F_o peaks, or for all high intensity portions or high energy integral portions of the speech. These and other further studies in stressed syllable location and constituent boundary detection will be outlined in section 6.

6. CONCLUSIONS AND FURTHER WORK

In this report, methods have been described for segmenting speech into grammatical phrases and identifying stressed syllables in continuous speech. The program for detecting syntactic boundaries from fall-rise patterns in voice fundamental frequency contours has been shown, both by the present study and by previous studies, to succeed in finding over 80% of all syntactically predicted boundaries between major syntactic units. It also, however, detects some syntactic boundaries not predicted by the intuitive constituent structure analysis previously applied, and detects false boundaries not apparently related to syntactic structure, such as at consonant-vowel boundaries.

The algorithm for stressed syllable location has succeeded in locating around 85% of all syllables perceived as stressed by the majority votes of a panel of listeners. The procedure identifies stressed syllables with high energy-integral portions of the speech which exhibit rising or non-falling F_o, but it does so in a way which makes use of constituent boundaries and archetype F_o contours. Simpler procedures might conceivably work as well, and there is obviously room for improvement in the present location scores.

Besides such algorithmic results, the other major aspect of research reported herein has been concerned with the perceptions of stress levels by three listeners. Two listeners were found to agree in their perceived stress levels for most of the individual syllables in the Rainbow Script and Monosyllabic Script, and ARPA man-machine interaction sentences. They differed on only about 5% of all syllables as to whether they were stressed or not, and each of them showed only about 5% confusions in decisions about stressed syllables from one trial to another. Unstressed and reduced levels were much more frequently confused. A third listener differed from the other two listeners on about half of his stress level judgments. About 20 to 25% of all syllables were labelled stressed by the other listeners, but unstressed by this third listener. This listener also labelled substantial

percentages of all syllables as stressed on one trial and unstressed on another. Such listeners who are inconsistent in their own judgments and who differ dramatically from other listeners should be excluded in any attempts to establish standards about which are the actual "stressed syllables" in connected speech.

The listeners appear to be as consistent in their assignments of stress levels given only the written text as they are in their assignments when listening to the speech recordings. However, their judgments without speech do not correspond well with their judgments with speech if the speech is spontaneous (that is, not produced by speakers reading written texts). Listeners apparently differ most dramatically from each other, and yield more confusions in stress levels from repetition to repetition, when yes-no questions are involved.

The majority stress perceptions from three trials by each listener, when pooled so as to yield the sum plots as shown in Figure 2, provide a "standard" for determining all stressed syllables which is stable to within about 5%. This is suitable for evaluating an algorithm for locating stressed syllables to within a 5% tolerance in overall location scores.

Several forms of further work are needed. The program for constituent boundary detection can be refined to produce fewer false alarms by requiring each new F_o maximum or minimum to remain beyond the 7% thresholds for at least 20 ms. It would be desirable to remove or augment the strict dependence on a fixed (7%) threshold for F_o changes, and to incorporate an overall confidence measure for each boundary, based on the percentage decrease in F_o before the apparent boundary, the percentage increase after the boundary, the shape of the contour near the boundary, and the time between that boundary and the immediately preceding or following ones. Thus, cusp-like changes at unvoiced consonants (of the form — ⌣) and very brief F_o dips or jumps (of such forms as ⌝—) may be assigned very low likelihood of being boundaries, while major gradual changes (of the form ⌣) would be assigned higher confidence ratings. One or both of two boundaries separated by short times (in the order of 200 ms or less) might be considered suspect, and assigned a low confidence rating.

The boundary predictions should be improved by defining and applying a strict set of rules for syntactic bracketing and prediction of intonation contours. Intonation rules such as Bierwisch (1966) produced for German are needed, along with the selection of an adequate grammar to define the syntactic structure that would be part of the input to such intonation rules. Working with Jane Robinson from the University of Michigan, we hope to apply such rules to texts such as those used in the present studies.

The algorithm for locating stressed syllables must be implemented as a computer program and tested carefully to see that it performs at the level of success attained in the previous hand analyses. Also, several improvements are needed. Among those to be investigated are better procedures for defining the TAIL of a constituent, a careful "tuning" of all the parameters and detailed steps for selecting HEADS and other stressed syllables, use of a low-frequency "sonorant" energy function rather than the present broadband energy function (so that better syllabication might be attained), and the incorporation of procedures for locating other possible stressed syllables before the HEAD (or peak F_o position) when the peak F_o occurs late in a constituent (say more than 400 or 500 ms after the preceding boundary).

It also seems reasonable to compare the results with the present stressed syllable algorithm (either before or after it is implemented as a computer program) with results in stressed syllable location by other possible procedures. For example, if one called all long-duration portions where energy was above a threshold value as stressed syllables, how many of the perceived stressed syllables would be detected and how many false alarms would result? Alternatively, could one get comparable success by looking for all F_o rises or upward inflections and choosing the high energy portion nearest such places, without use of boundaries and archetype contours in his procedures?

More extensive experiments are needed wherein the various variables of sentence type, talker, lexical forms, phonetic content, position in sentence and intonation contour, and such could be independently controlled. Texts for such studies are now being designed (cf. Lea, Medress, and Skinner, 1972a,

pp. 56-57), and such studies will be conducted. In particular, such studies can test further the apparent difficulty in listeners' assignments of stress within yes-no questions, and the relative successes in boundary detection and stressed syllable location within questions versus declaratives or commands.

The application of boundary detections and stressed syllable locations to guiding a partial distinctive features analysis must yet be done. Until some details of the distinctive features analysis are better defined, the question cannot be resolved as to whether higher "hit" rates or lower "false alarm" rates are more important to attain in the boundary detection or stressed syllable location algorithm. Also, techniques must be explored for applying boundary and stressed syllable information to the aid of syntactic parsers. Such efforts will be critical to implementing the proposed speech recognition strategy at Univac.

7. REFERENCES

ARMSTRONG, L. E. and WARD, I. C. (1926), Handbook of English Intonation. Cambridge: Heffer (2nd Edit.).

BIERWISCH, M. (1965), Regeln fur die Intonation deutscher Satze, Studia Grammatica, vol. 7, pp. 99-201.

BLACKMAN, R. B., and TUKEY, J. W., (1958) The Measurement of Power Spectra, Dover Publication Inc., New York.

BOLINGER, D. (1958), A Theory of Pitch Accent in English. Word, vol. 14, p. 109.

CHOMSKY, N. (1965), Aspects of the Theory of Syntax. Cambridge, Mass: M.I.T. Press.

CHOMSKY, N. and HALLE, M. (1968), The Sound Pattern of English. New York: Harper and Row.

EMONDS, J. E. (1970), Root and Structure Preserving Transformations, Ph. D. Thesis, Linguistics Dept., M.I.T.

FAIRBANKS, G. (1940), Voice and Articulation Drillbook. New York: Harper and Row.

FRY, D. B. (1955), Duration and Intensity as Physical Correlates of Linguistic Stress. J. Acoust. Soc. Amer., vol. 35, pp. 765-769.

FRY D. B. (1958), Experiments in the Perception of Stress. Language and Speech, vol. 1, pp. 126-152.

GOLDMAN-EISLER, F. (1961), A Comparative Study of Two Hesitation Phenomena, Language and Speech, vol. 4, pp. 18-26.

HALLE, M. and KEYSER, S. J. (1971), English Stress. New York: Harper and Row.

HOUSE, A. S. and FAIRBANKS, G. (1953), The Influence of Consonant Environment upon the Secondary Acoustical Characteristics of Vowels, J. Acoust. Soc. Amer., vol. 25, pp. 105-113.

JONES, D. (1932), Outline of English Phonetics. Cambridge: Heffer (8th Ed.).

LEA, W. A. (1971), A Formalization of Measurement Scale Forms. Journal of Math. Sociology, Vol. 1, 81-104.

LEA, W. A. (1972a), An Approach to Syntactic Recognition without Phonemics. Proc. 1972 Intern. Conf. on Speech Commun. and Processing. Newton, Mass.: pp. 198-201.

LEA, W. A. (1972b), Intonational Cues to the Constituent Structure and Phonemics of Spoken English, Ph.D. Thesis, School of E.E., Purdue University.

LEA, W. A. (1973a), Segmental and Suprasegmental Influences on Fundamental Frequency Contours. Presented at the Symposium on Consonant Types and Tone, University of Southern California, Los Angeles, March 9-10, 1973. To appear in Consonant Types and Tone (Proceedings of the First Annual Southern California Round Table in Linguistics, Ed. by L. Hyman), University of Southern California.

LEA, W. A. (1973b), Use of Prosodic Features to Segment Continuous Speech into Sentences and Phrases, Univac Report PX 10058. To be published, under the title "An Approach to Syntactic Recognition Without Phonemics", in IEEE Trans. on Audio and Electroacoustics, Vol. AU-21, No. 3, June, 1973.

LEA, W. A., MEDRESS, M. F., and SKINNER, T. E. (1972a), Prosodic Aids to Speech Recognition I: Basic Algorithms and Stress Studies, Univac Report No. PX 7940, Univac Park, St. Paul, Minnesota.

LEA, W. A., MEDRESS, M. F., and SKINNER, T. E. (1972b), Use of Syntactic Segmentation and Stressed Syllable Location in Phonemic Recognition. Presented at the 84th Meeting, Acoustical Society of America, Miami Beach, Florida, Nov. 27-30, 1972.

LEHISTE, I. (1970), Suprasegmentals. Cambridge: M.I.T. Press.

LI, K.-P., HUGHES, G. W., and SNOW, T. B. (1973), Segment Classification in Continuous Speech, IEEE Trans. on Audio and Electracoustics, Vol. AU-21, No. 1, pp. 50-57.

LIEBERMAN, P. (1960), Some Acoustic Correlates of Word Stress in American English. J. Acoust. Soc. Amer., vol. 32, pp. 451-454.

LIEBERMAN, P. (1967), Intonation, Perception, and Language. Cambridge: M.I.T. Press.

MAKHOUL, J. (1972), Aspects of Linear Prediction in the Spectral Analysis of Speech, Proc. 1972 Conf. on Speech Communication and Processing, Newton, Mass., pp. 77-81.

MATTINGLY, I. (1966), Synthesis by Rule of Prosodic Features. Lang. and Speech, vol. 9, pp. 1-13.

MEDRESS, M. F., and SKINNER, T. E., and ANDERSON, D. E. (1971), Acoustic Correlates of Word Stress, Presented to 82nd Meeting, Acoustical Society of America, Denver, Colorado, October 20, (Paper K3).

MORTON, J. and JASSEM, W. (1965), Acoustic Correlates of Stress, Lang. and Speech, vol. 8, 159-81.

NEWELL, A., et. al (1971), Speech-Understanding Systems: Final Report of a Study Group. Pittsburgh, Penn.: Carnegie-Mellon University.

OLLER, D. K. (1971), The Effect of Position-in-Utterance and Word-Length on Speech Segment Duration. Unpublished manuscript, Depts. of Psychology and Linguistics, University of Texas, Austin.

PIKE, K. L. (1945), The Intonation of American English. Ann Arbor: University of Michigan.

SAUSSURE, F. (1915), Course in General Linguistics, (translated in 1959 from the 1915 materials, by W. Baskin). New York: The Philosophical Library.

SONDHI, M. M. (1968), New Methods of Pitch Extraction, IEEE Trans. on Audio and Electroacoustics, vol. AU-16, pp. 262-266.

STEVENS, S. S. (1951), Mathematics, Measurement, and Psychophysics. In Handbook of Experimental Psychology (S. S. Stevens, Ed.). New York: John Wiley and Sons, 1951, 1-49.

WALKER, D. E. (1973), Speech Understanding Research, Annual Technical Report prepared for ARPA, Contract DAHC04-72-C-0009, Stanford Research Institute, Menlo Park, California.

WOODS, W. A. (1971), The Lunar Sciences Natural Language Information System, BBN Report No. 2265, NASA Contract NAS9-1115, Bolt Beranek and Newman, Cambridge, Massachusetts.

APPENDIX A: CONSTITUENT BOUNDARY DETECTION RESULTS

The constituent boundary detection program marks boundaries between major syntactic units by locating the last time of minimum F_0 value, in an F_0 "valley" which is preceded by a 7% decrease in F_0 and followed by a 7% increase. A general flow chart of the procedure was published in Lea's thesis (Lea, 1972b, p. 206). A detailed flowchart (available upon request) has been obtained by an automatic flow-charting routine at Sperry Univac, for that version implemented at Univac and used for boundary detection on the Monosyllabic Script and the ARPA Sentences.

Figures A-1, A-2, and A-3 show the detected boundaries for the Rainbow Script (as spoken by six talkers), the Monosyllabic Script (as spoken by two talkers), and the ARPA Sentences, respectively. Vertical bars mark predicted constituent boundaries that were detected; predicted boundaries that were not detected (that is, were "missing") are indicated by asterisks at the positions of the syntactic breaks. Boundaries between minor syntactic constituents (that were detected but not predicted) are shown by columns of dots, and false (syntactically unrelated) boundaries that were detected are shown by question marks. Sentence boundaries were expected to be accompanied by both the vertical bars marking F_0 - valley constituent boundaries and by pauses of 35 centiseconds or more, to be marked by S's on the vertical bars. When a sentence boundary was not accompanied by a sufficient pause, but was detected as a constituent boundary, it was marked by this symbol: ⋮ . In the ARPA Sentences, occasional extra hesitation pauses occur that are not associated with major syntactic boundaries. These are marked in Figure A-3 by S's with columns of dots (not vertical bars).

Table A-1 shows the boundary detection results for the 13 ARPA Sentences, separated into categories for each type of sentence. WH-questions and commands show the most missing, or undetected, boundaries, thus yielding the lowest constituent boundary detection scores. Three of the missing boundaries in the commands, and two of those missing in WH-questions, are in compound noun constructions, which are certainly among the most minor of the predicted boundaries. Another missing boundary in a command is a

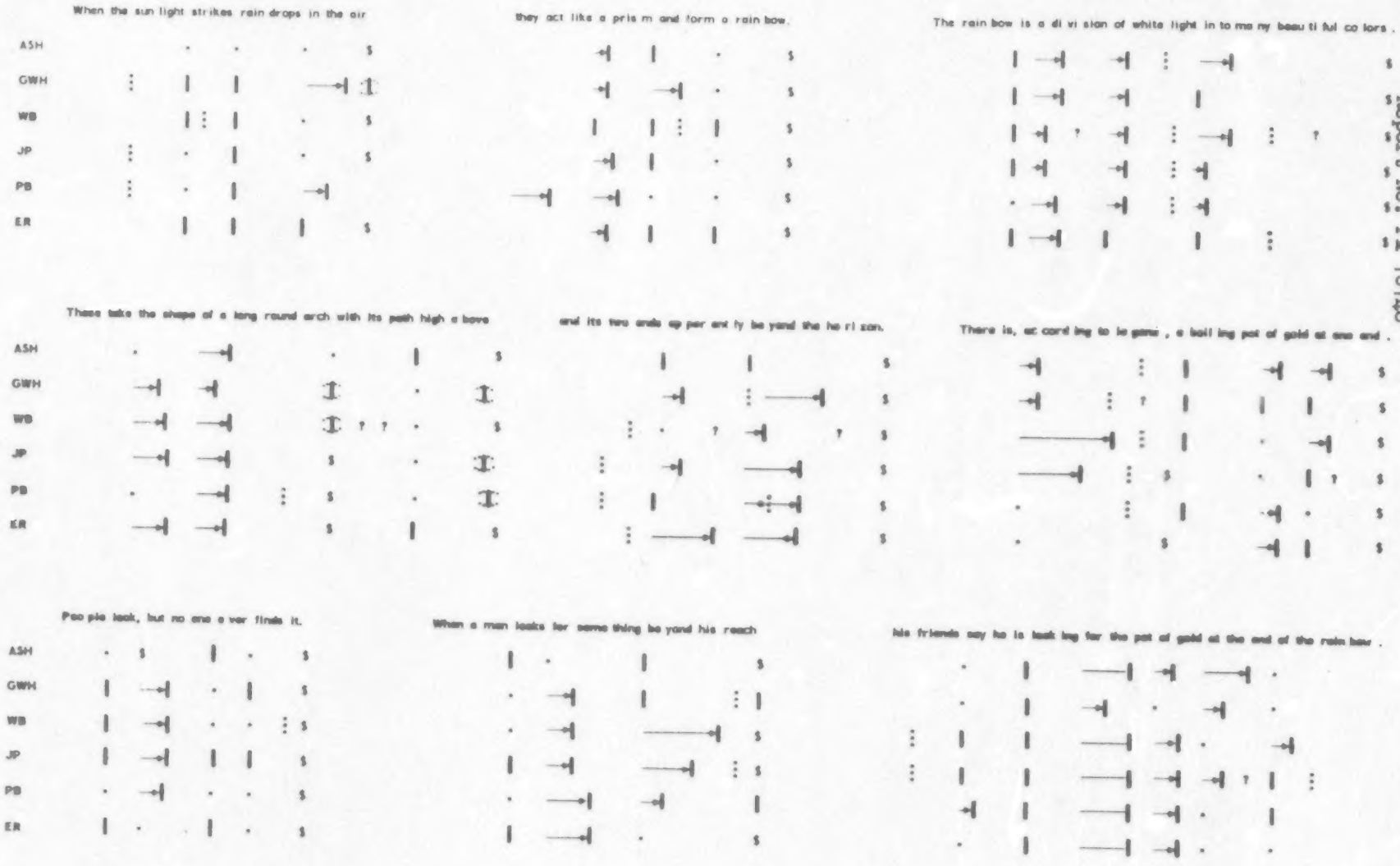

When the sun light strikes rain drops in the air
they act like a pris m and form a rain bow.
The rain bow is a di vi sion of white light in to ma ny beau ti ful co lors .
These take the shape of a long round arch with its path high a bove
and its two ends ap pear ex ly be yond the ho ri zon.
There is, ac cord ing to le gend , a boil ing pot of gold at one end .
Peo ple look, but no one e ver finds it.
When a man looks for some thing be yond his reach
his friends say he is look ing for the pot of gold at the end of the rain bow .
ASH
GWH
WB
JP
PB
ER

Figure A-1. Complete Boundary Detection Results for the Rainbow Script Spoken by Six Talkers. Predicted boundaries that were detected are shown by vertical bars (|); boundaries between minor syntactic constituents (that were detected but not predicted) are shown by columns of dots (⋮); and false (syntactically unrelated) boundaries that were detected are shown by question marks (?). Asterisks (*) are marked at positions where syntactic boundaries had been predicted but had not been detected by the boundary detection algorithm. Sentence boundaries that were accompanied by 35 csec pauses are shown by $'s with the vertical bars of constituent boundaries ($); a sentence boundary which was not accompanied by a pause, but was detected as a constituent boundary is shown without the $, as follows (‡).

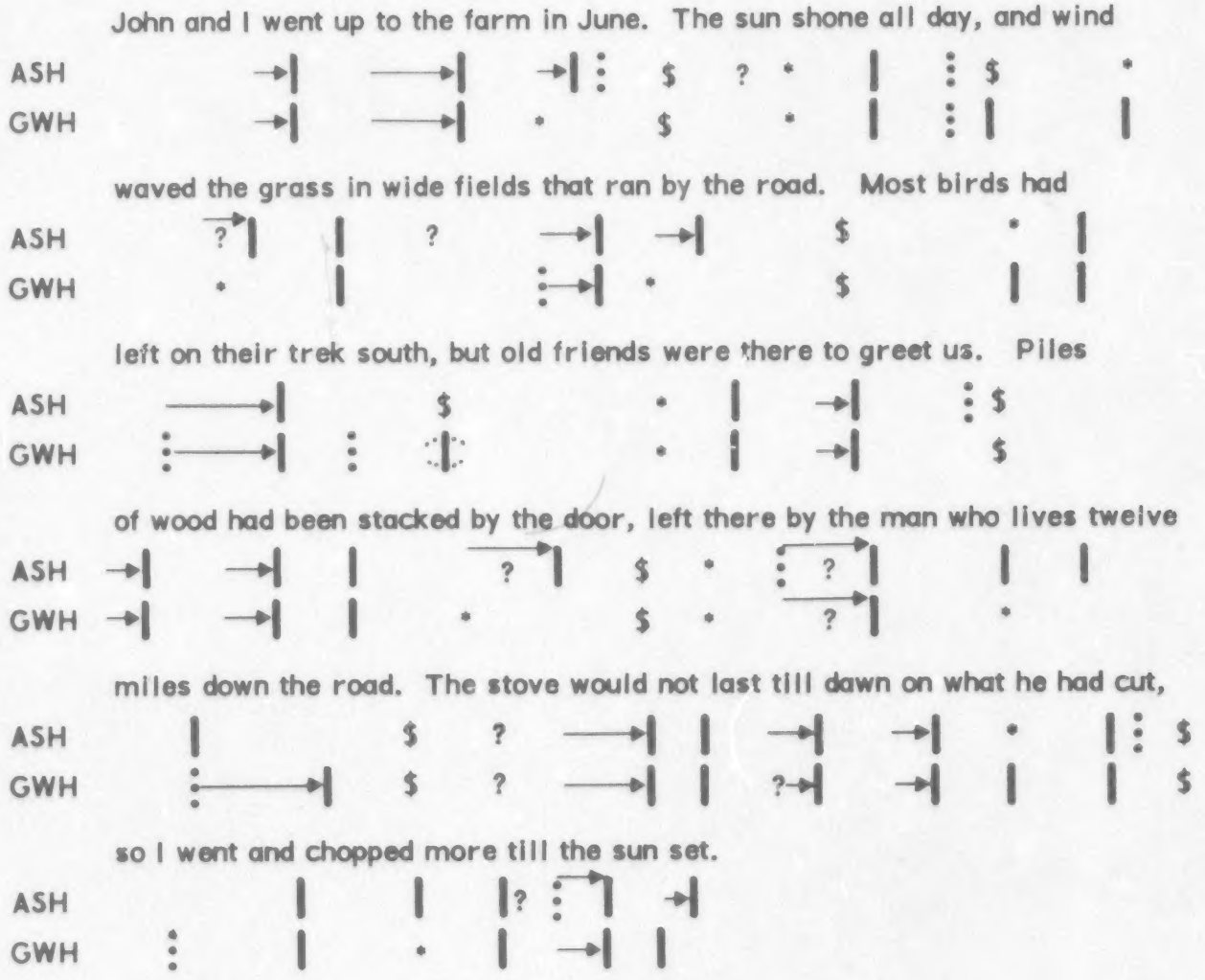
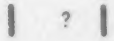

John and I went up to the farm in June. The sun shone all day, and wind
ASH
GWH
waved the grass in wide fields that ran by the road. Most birds had
ASH
GWH
left on their trek south, but old friends were there to greet us. Piles
ASH
GWH
of wood had been stacked by the door, left there by the man who lives twelve
ASH
GWH
miles down the road. The stove would not last till dawn on what he had cut,
ASH
GWH
so I went and chopped more till the sun set.
ASH
GWH

Figure A-2. Complete Boundary Detection Results for the Monosyllabic Script for Talkers ASH and GWH. Boundary detection symbols are explained in Figure A-1.

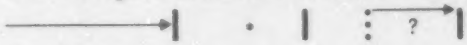
1. (LS21) Who's the owner of utterance eight?

2. (LM13) Display the phonemic labels above the spectrogram.

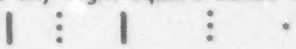
3. (B27) Do any samples contain troilite?



4. (B10) What is the average uranium lead ratio for the lunar samples?



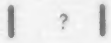
5. (RB6) Do you have any right square boxes left?



6. (RB16) Put the other red block on the red block.

7. (LM3) Who is the owner of utterance eight?

8. (B35) Do any samples contain tridymite?



9. (RA19) Would you move the stack of right circular cylinders to the right by half a square?

10. (RC8) Place the red triangle two squares back from the front of the floor in the middle.

11. (CB1300) Alpha becomes alpha minus beta.

12. (CB2300) Alpha gets alpha minus beta.

13. (D10) Repeat where key word equals Gauss elimination or key word equals eigenvalues.

Figure A-3. Complete Boundary Detection Results for the 13 ARPA Sentences. Symbols marking boundaries are explained in Figure A-1.

TABLE A-I.

BOUNDARY DETECTION RESULTS FOR VARIOUS SENTENCE TYPES
(DATA FROM 13 ARPA SENTENCES)

SENTENCE TYPE	NUMBER OF PREDICTED BOUNDARIES CORRECTLY DETECTED	NUMBER OF PREDICTED BOUNDARIES MISSED (I.E., NOT DETECTED)	NUMBER OF "EXTRA" BOUNDARIES DETECTED	NUMBER OF "FALSE" BOUNDARIES DETECTED
YES-NO QUESTIONS	6	1	2	2
WH QUESTIONS	5	5	1	1
COMMANDS	14	5	4	2+4 Pauses
DECLARATIVES	6	1	0	1
POLITE COMMANDS	5	1	2	1+4 Pauses

NP-Verbal boundary, which has been shown to be a type of boundary which is frequently missing. Boundaries are also missing after the WH-pronoun plus copulatives ("Who's" or "Who is"). Boundaries might be argued to be less likely there anyway, since previous results have shown that pronouns and copulatives both are less likely to be followed by detectable boundaries.

The only boundaries in WH-questions and commands that are notable in their absence, then, are that after the command verb Display in LM13 and that before the preposition phrase of LM3. These misses are apparently due to the montonic speech of that particular talker.

We are left with little or no evidence that sentence type affects relative boundary detection scores, except for the WH-pronoun effects.

The extra pauses in the command and polite command are not necessarily results of sentence type, but are all hesitation pauses in the spontaneous protocols from Stanford Research Institute.

Since boundary detection scores were somewhat lower in the ARPA Sentences, and since the monotonic F_0 patterns in those spontaneous utterances seemed to be one factor in the results, a study was conducted on the thresholds for detecting fall-rise valleys in F_0, and how they correlate with boundary scores, for the 6ARPA sentences. In Figure A-4 is shown the number of predicted, extra, and false boundaries detected in the 6ARPA sentences as a function of threshold. As the threshold is increased (that is, a boundary must be preceded by larger F_0 decreases and followed by larger F_0 increases), the number of false boundaries rapidly drops while the numbers of predicted and extra syntactic boundaries decreases much more gradually. Any threshold above 3% and below 10% or so thus eliminates most false boundaries while preserving the detection of most predicted boundaries.

The threshold plotted along the abscissa in Figure A-4 is the smaller of the two thresholds for percentage rise in F_0 and percentage fall in F_0. Past work (Lea, 1972b) has been conducted with both thresholds equal. The Univac implementation permits unequal fall and rise thresholds.

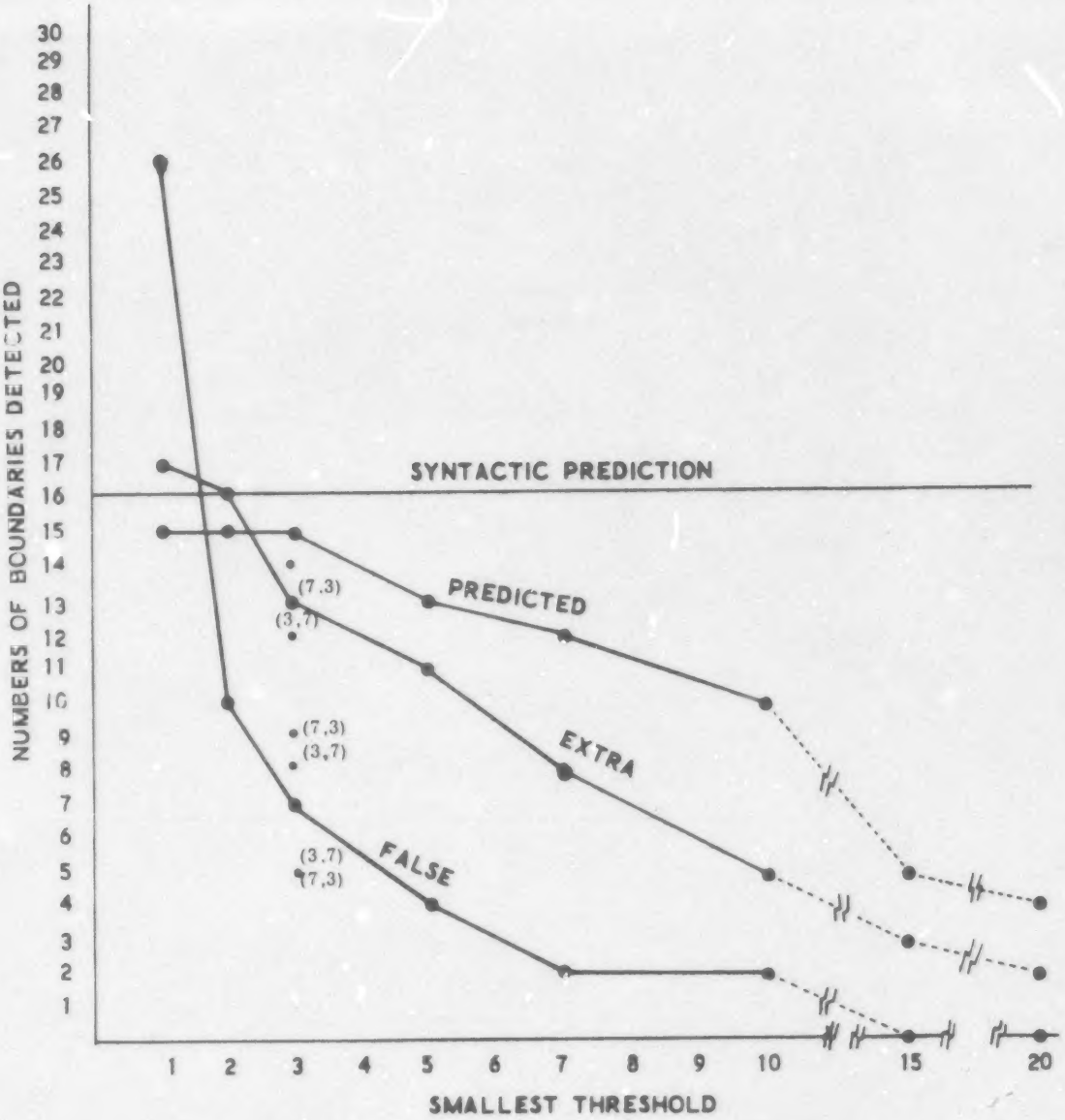


Figure A-4. Effects of Threshold Size on Boundary Detection Results, for the 6 ARPA Sentences.

Figure A-4 shows some results with unequal thresholds. Plotted at the smallest-threshold value of 3% are:

(a) the results with a minimum fall of 7% required while only 3% following rise is required, symbolized by the pair (7,3); and

(b) the results with a minimum fall of only 3% required while a 7% following rise is required, symbolized by the pair (3,7).

Thus, when the fall threshold (the first in the pair) is greater than the rise threshold, more predicted and extra syntactic boundaries are correctly detected, and less false boundaries are detected, than if the rise threshold were greater than the fall threshold. Of course, fewer boundaries of all types are detected if both thresholds are increased, but for nonequal thresholds, the fall threshold should be greater than the rise threshold. This is to be expected when one considers the general falling contours of F_0 or intonation in English (see figures 9, 10, and 12 of this report).

APPENDIX B: DETAILS OF PERCEIVED STRESS PATTERNS

Figure B-1 illustrates a sheet on which the perceived stress levels of one listener are recorded for one recorded text, the 6ARPA Sentences. Similar sheets were obtained for each trial with each listener, each text, and each talker. Stressed, unstressed, and reduced syllables were marked as S, U, ard R, respectively, by this listener (MFM) and another listener (WAL). Listener TES labelled them as levels 1, 2, and 3, respectively. Vertical lines delimited syllables, to facilitate marking for every syllable.

Figures B-2 to B-15 summarize the majority perceptions from three repetitions for three listeners. The majority perceptions for each listener were first obtained (for each text and talker) from three repetitions. Then the number of majority votes of a syllable as stressed, minus the number of votes as reduced, were plotted under each syllable ("unstressed" judgments were thus assigned zeros, neither adding to nor subtracting from the syllable's stress score). Figures B-2 to B-8 show the results for the Rainbow Script spoken by six talkers and for the NO SPEECH condition where only the written text was provided to the subjects. Figures B-9 to B-11 show results for the Monosyllabic Script with two talkers and NO SPEECH conditions. Figures B-12 and B-13 are corresponding SPEECH and NO SPEECH results for 6ARPA, while B-14 and B-15 are for 7ARPA.

STRESS PERCEPTIONS ON ARPA SENTENCES

Listener MFM Date 2/20/73

LS21:

S	U	R	S	U	R	S	R	U	S
Who's	the	owner	of	utterance					eight?

LM13:

U	S	U	U	S	U	S	U	U	S	U	S	U	U
Display	the	phonemic				labels			above	the	spectrogram.		

B27:

U	U	R	S	U	U	S	S	U
Do	any	samples			contain		troilite?	

B10:

S	U	R	S	R	U	U	S	U	U	S	S	U	U	U	R	S	R	S	U
What	is	the	av	er	age				uranium	lead	ra	tio		for	the	lu	nar	samples?	

RB6:

S	S	S	U	R	S	S	S	R	U
Do	you	have	any		right	square		boxes	left?

RB16:

S	U	S	U	S	S	S	S	U	S	S
Put	the	other		red	block	on	the	red	block.	

Figure B-1. Sample of the Sheets Used for Marking Stress Judgments (Listener MFM)

WHEN THE SUN LIGHT STRIKES RAIN DROPS IN THE AIR, THEY ACT LIKE A PRIS M AND FORM A RAIN BOW. THE RAIN BOW IS A DI VI SION OF WHITE LIGHT IN TO MA NY BEAU TI FUL CO LORS. THESE TAKE THE SHAPE OF A LONG ROUND ARCH WITH ITS PATH HIGH A BOVE AND ITS TWO ENDS AP PAR ENT LY
+3 +2 +1 0 -1 -2 -3
BE YOND THE HO RI ZON. THERE IS AC CORD ING TO LE GEND A BOI LING POT OF GOLD AT ONE END. PLE NTY OF GOLD, BUT NO ONE EV ER FINDS IT. WHEN A MAN LOOKS FOR SOME THING BE YOND HIS REACH HIS FRIENDS SAY HE IS LOOK ING FOR THE POT OF GOLD AT THE END OF THE RAIN BOW
+3 +2 +1 0 -1 -2 -3

Figure B2. Summary of Stress Judgments by Three Listeners, for One Talker (ASH) Reading the Rainbow Script. Plotted for each syllable is the number of listeners whose majority judgments (from three trials) declare the syllable as stressed minus the number of such majority judgments of the syllable as reduced. Unanimous judgment as stressed thus yields the top value of +3, whereas judgments as reduced pull the value down toward -3.

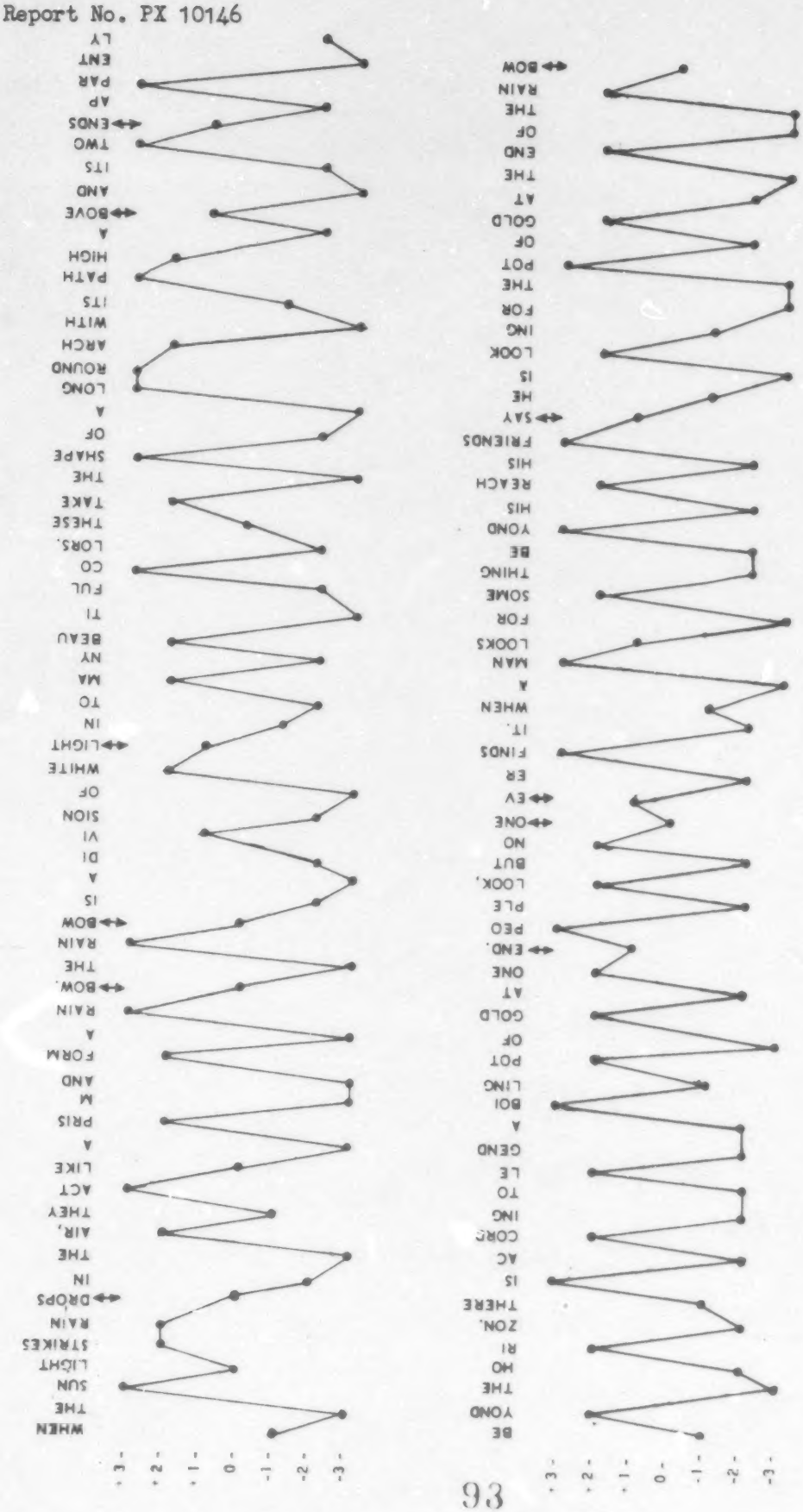


Figure B-3. Summary of Stress Judgments by Three Listeners, for Talker GWH Reading the Rainbow Script

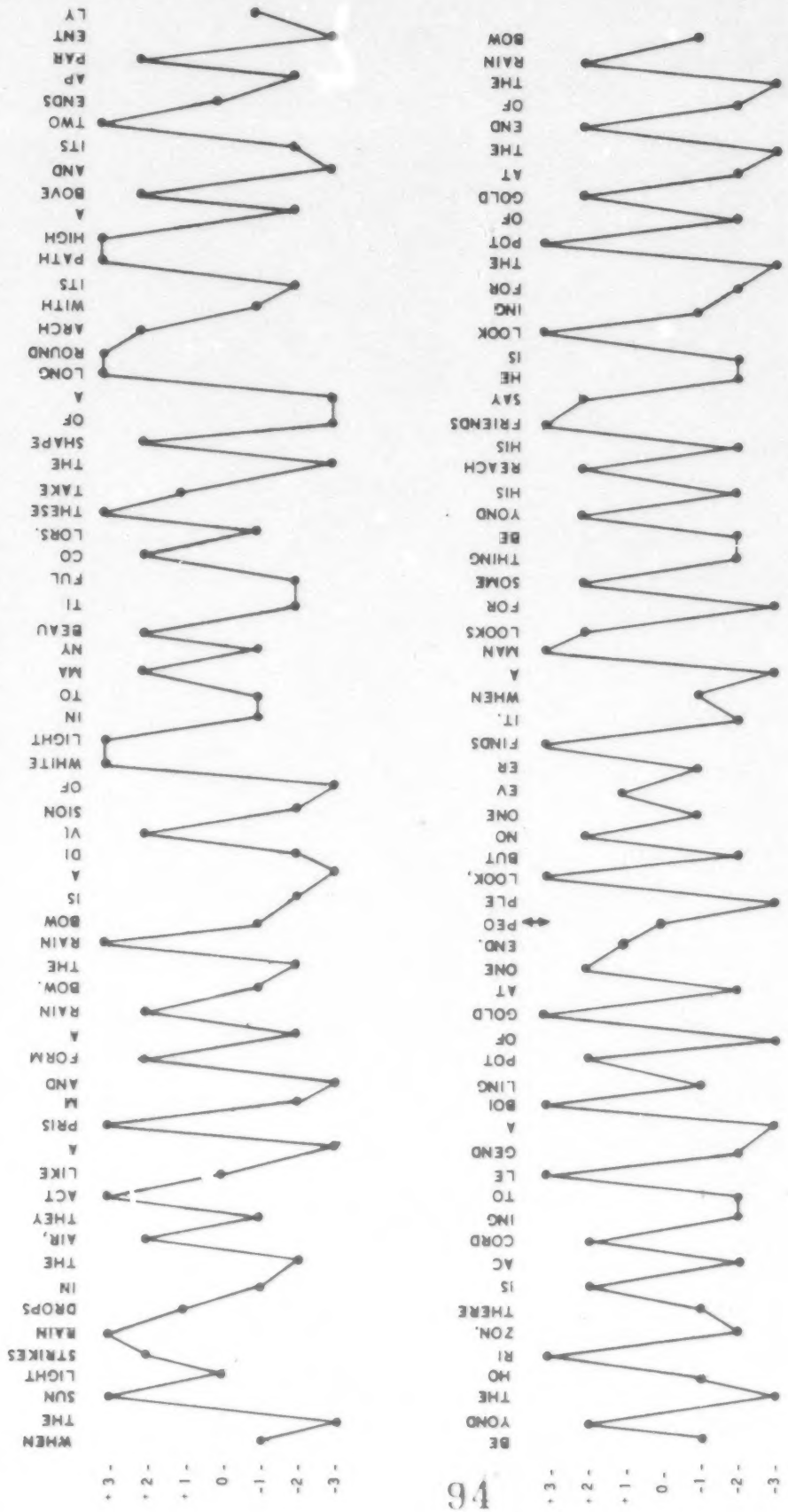

+3
+2
+1
0
-1
-2
-3
WHEN THE SUN LIGHT STRIKES RAIN DROPS IN THE AIR, THEY ACT LIKE A PRIS M AND FORM A RAIN BOW. THE RAIN BOW IS A DI VI SION OF WHITE LIGHT IN TO MA NY BEAU TI FUL CO LORS. THESE TAKE THE SHAPE OF A LONG ROUND ARCH WITH ITS PATH HIGH A BOVE AND ITS TWO ENDS AP PAR ENT LY
BE YOND THE HO RI ZON. THERE IS AC CORD ING TO LE GEND A BOI LING POT OF GOLD AT ONE END. PEO PLE LOOK, BUT NO ONE EV ER FINDS IT. WHEN A MAN LOOKS FOR SOME THING BE YOND HIS REACH HIS FRIENDS SAY HE IS LOOK ING FOR THE POT OF GOLD AT THE END OF THE RAIN BOW

Figure B-4. Summary of Stress Judgments by Three Listeners, for Talker WB Reading the Rainbow Script

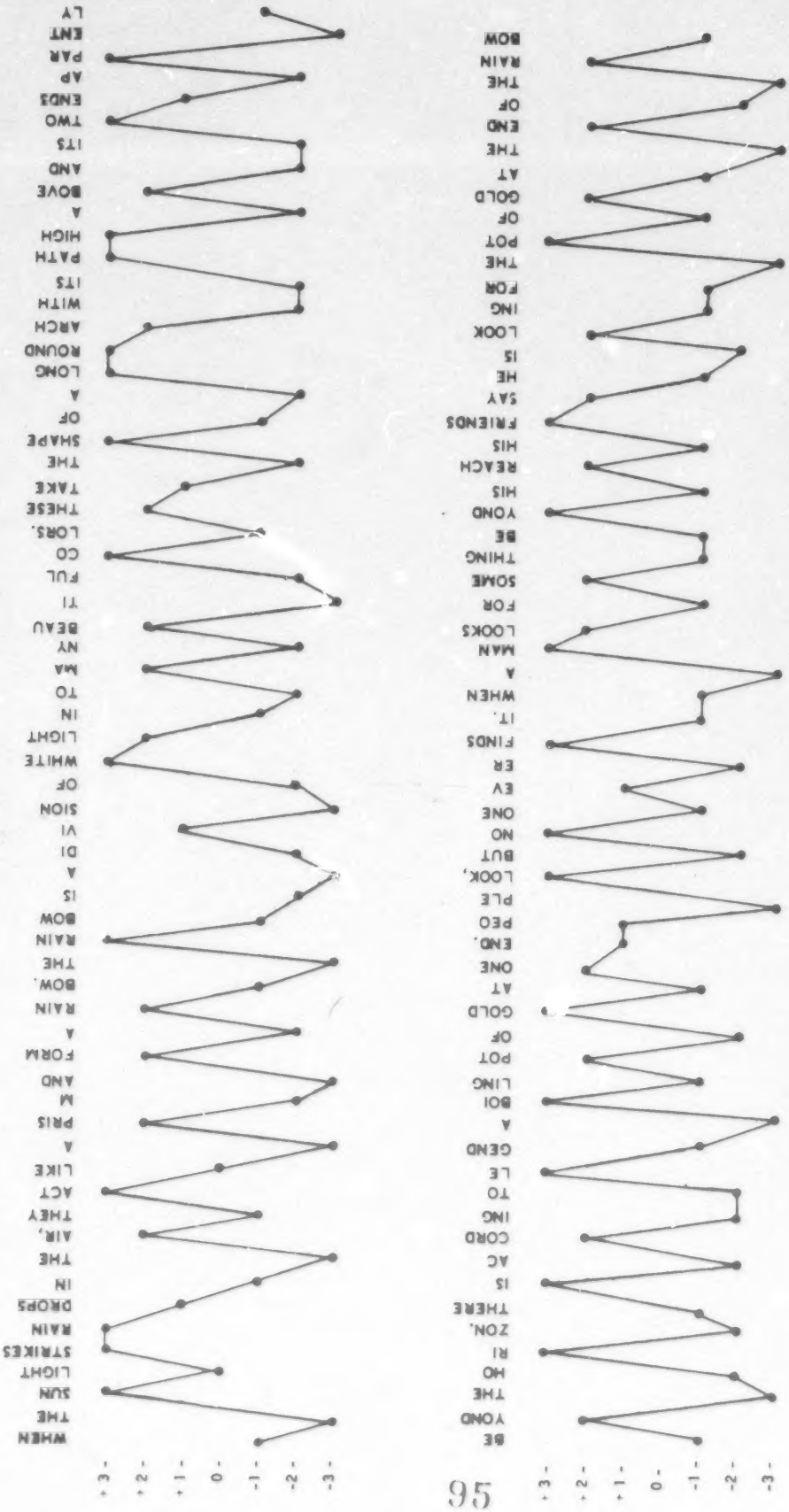

WHEN THE SUN LIGHT STRIKES RAIN DROPS IN THE AIR, THEY ACT LIKE A PRIS M AND FORM A RAIN BOW. THE RAIN BOW IS A DI VI SION OF WHITE LIGHT IN TO MA NY BEAU TI FUL CO LORS. THESE TAKE THE SHAPE OF A LONG ROUND ARCH WITH ITS PATH HIGH A BOVE AND ITS TWO ENDS AP PAR ENT LY
+3 +2 +1 0 -1 -2 -3
BE YOND THE HO RI ZON, THERE IS AC CORD ING TO LE GEND A BOI LING POT OF GOLD AT ONE END. PEO PLE LOOK, BUT NO ONE EV ER FINDS IT. WHEN A MAN LOOKS FOR SOME THING BE YOND HIS REACH HIS FRIENDS SAY HE IS LOOK ING FOR THE POT OF GOLD AT THE END OF THE RAIN BOW
+3 +2 +1 0 -1 -2 -3

Figure B-5. Summary of Stress Judgments by Three Listeners, for Talker JP Reading the Rainbow Script

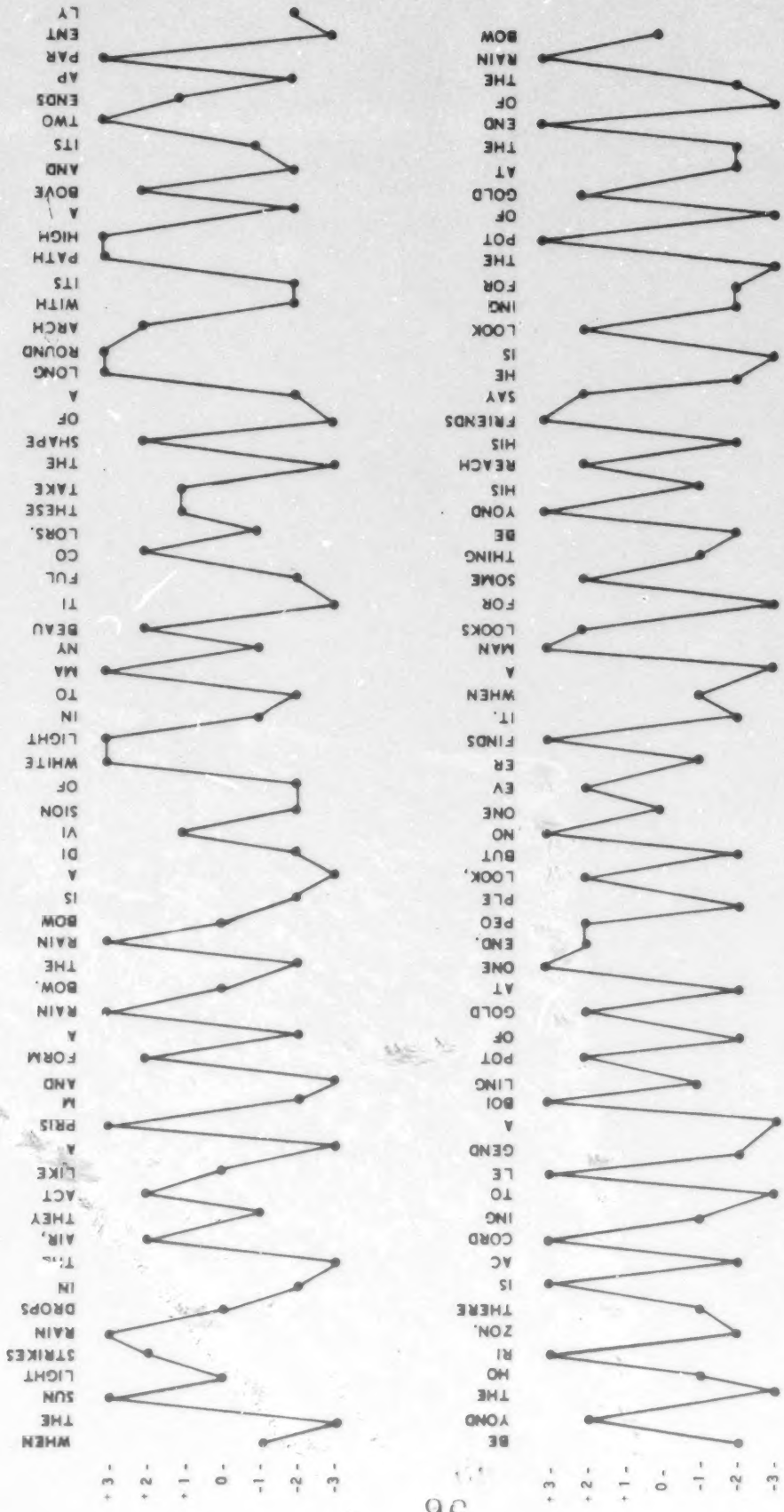


Figure B-6. Summary of Stress Judgments by Three Listeners, for Talker PB Reading the Rainbow Script

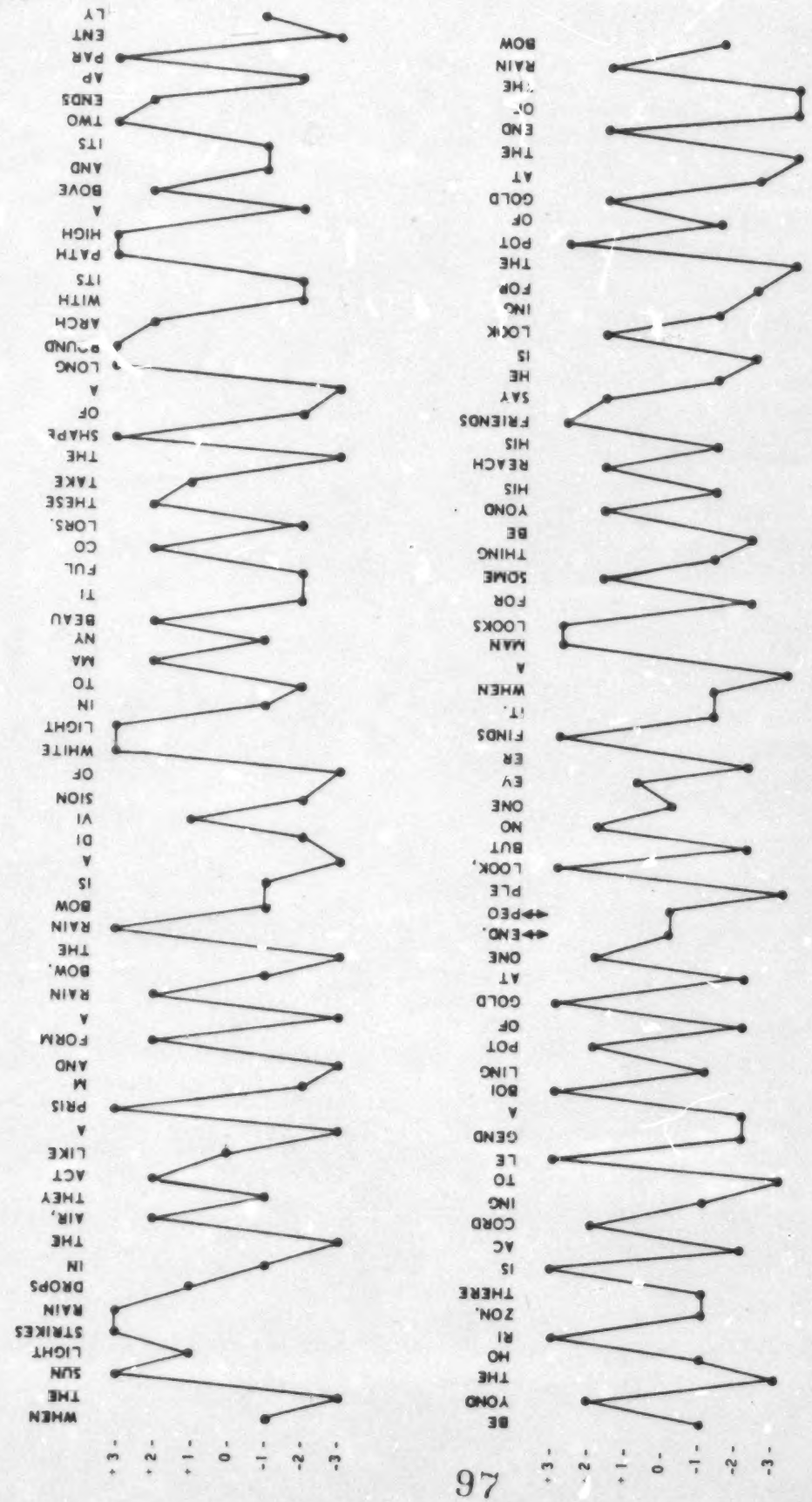

WHEN THE SUN LIGHT STRIKES RAIN DROPS IN THE AIR, THEY ACT LIKE A PRIS M AND FORM A RAIN BOW. THE RAIN BOW IS A DI VI SION OF WHITE LIGHT IN TO MA NY BEAU TI FUL CO LORS. THESE TAKE THE SHAPE OF A LONG ROUND ARCH WITH ITS PATH HIGH A BOVE AND ITS ENDS AP PAR ENT LY
+3 +2 +1 0 -1 -2 -3
BE YOND THE HO RI ZON. THERE IS AC CORD ING TO LE GEND A BOI LING POT OF GOLD AT ONE END. <-> PEO PLE LOOK, BUT NO ONE EV ER FINDS IT. WHEN A MAN LOOKS FOR SOME THING BE YOND HIS REACH HIS FRIENDS SAY HE IS LOCK ING FOR THE POT OF GOLD AT THE END OF THE RAIN BOW
+3 +2 +1 0 -1 -2 -3

Figure B-7. Summary of Stress Judgments by Three Listeners for Talker ER Reading the Rainbow Script.

Figure B-8. Summary of Stress Judgments by Three 'Listeners', When Given Only the Written Text of the Rainbow Script (NO SPEECH).

Figure B-9. Summary of Stress Judgments by Three Listeners, for Talker ASH Reading the Monosyllabic Script.

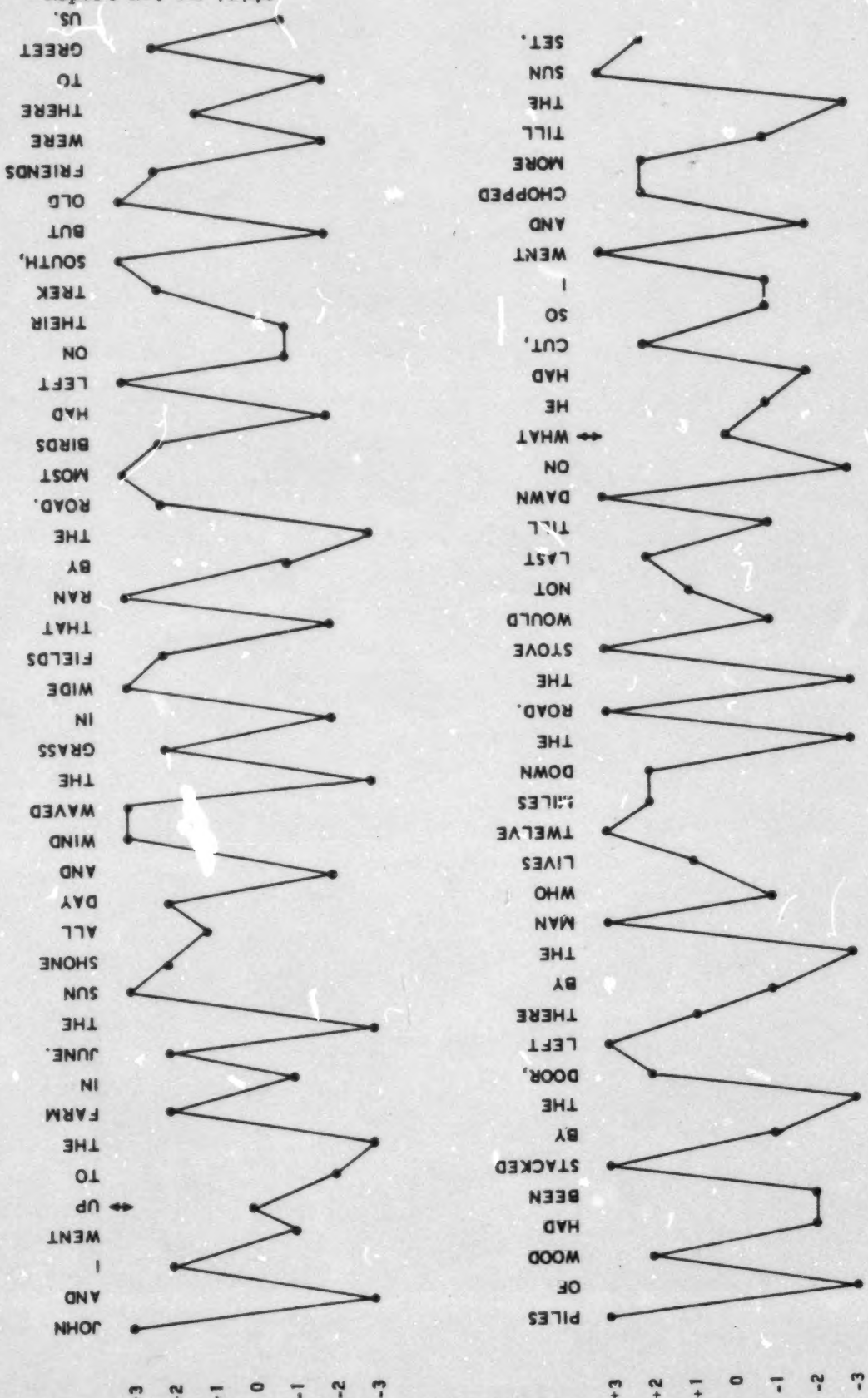


Figure B-10. Summary of Stress Judgments by Three Listeners, for Talker GWH Reading the Monosyllabic Script.

Figure B-11. Summary of Stress Judgments by Three 'Listeners', When Given Only the Written Text of the Monosyllabic Script (NO SPEECH).

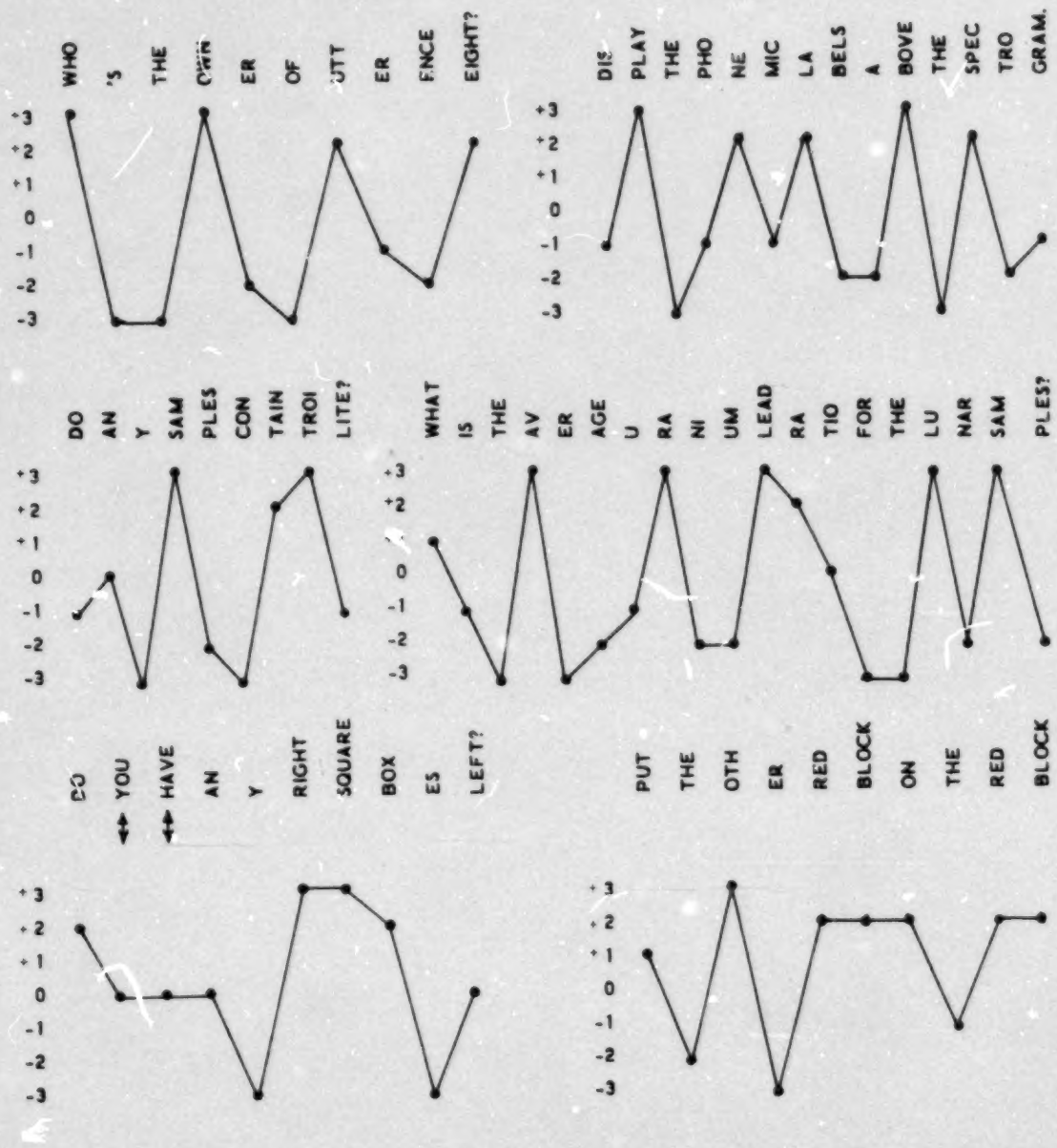

WHO 'S THE CHAIR OF UTTER FENCE EIGHT?
DIS PLAY THE PHO NE MIC LA BELS A BOVE THE SPEC TRO GRAM.
DO AN Y SAM PLES CON TAIN TROI LITE?
WHAT IS THE AV ER AGE U RA NI UM LEAD RA TIO FOR THE LU NAR SAM PLES?
DO YOU HAVE AN Y RIGHT SQUARE BOX ES LEFT?
PUT THE OTH ER RED BLOCK ON THE RED BLOCK
+3 +2 +1 0 -1 -2 -3

Figure B-12. Summary of Stress Judgments by Three Listeners, for the 6ARPA Sentences as Spoken.

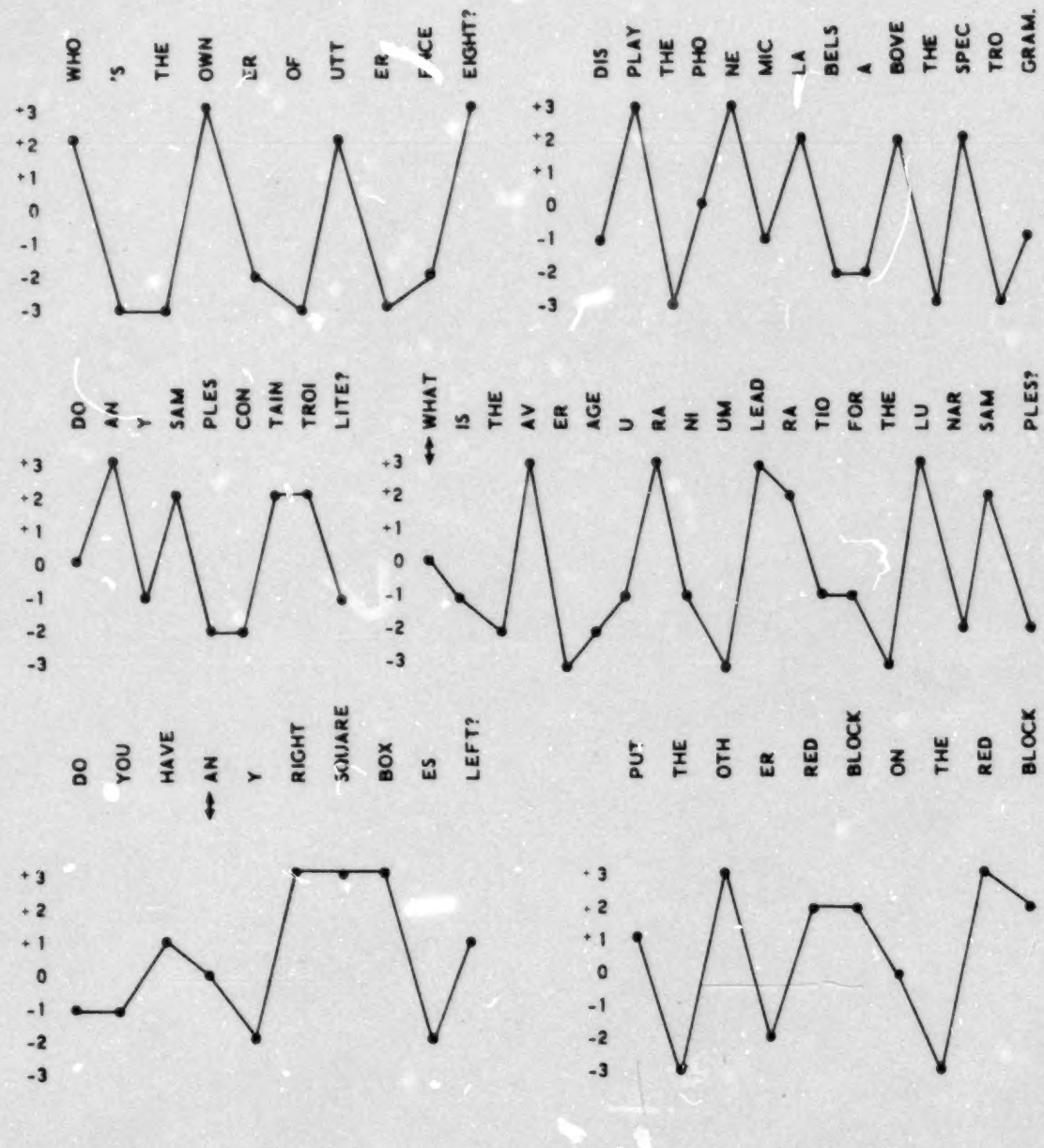

WHO 'S THE OWN ER OF UTT ER F ACE EIGHT?
DIS PLAY THE PHO NE MIC LA BELS A BOVE THE SPEC TRO GRAM.
DO AN Y SAM PLES CON TAIN TROI LITE?
WHAT IS THE AV ER AGE U RA NI UM LEAD RA TIO FOR THE LU NAR SAM PLES?
DO YOU HAVE AN Y RIGHT SQUARE BOX ES LEFT?
PUT THE OTH ER RED BLOCK ON THE RED BLOCK
+3
+2
+1
0
-1
-2
-3

Figure B-13. Summary of Stress Judgments by Three 'Listeners', When Given Only the Written Text of the 6ARPA Sentences (NO SPEECH).

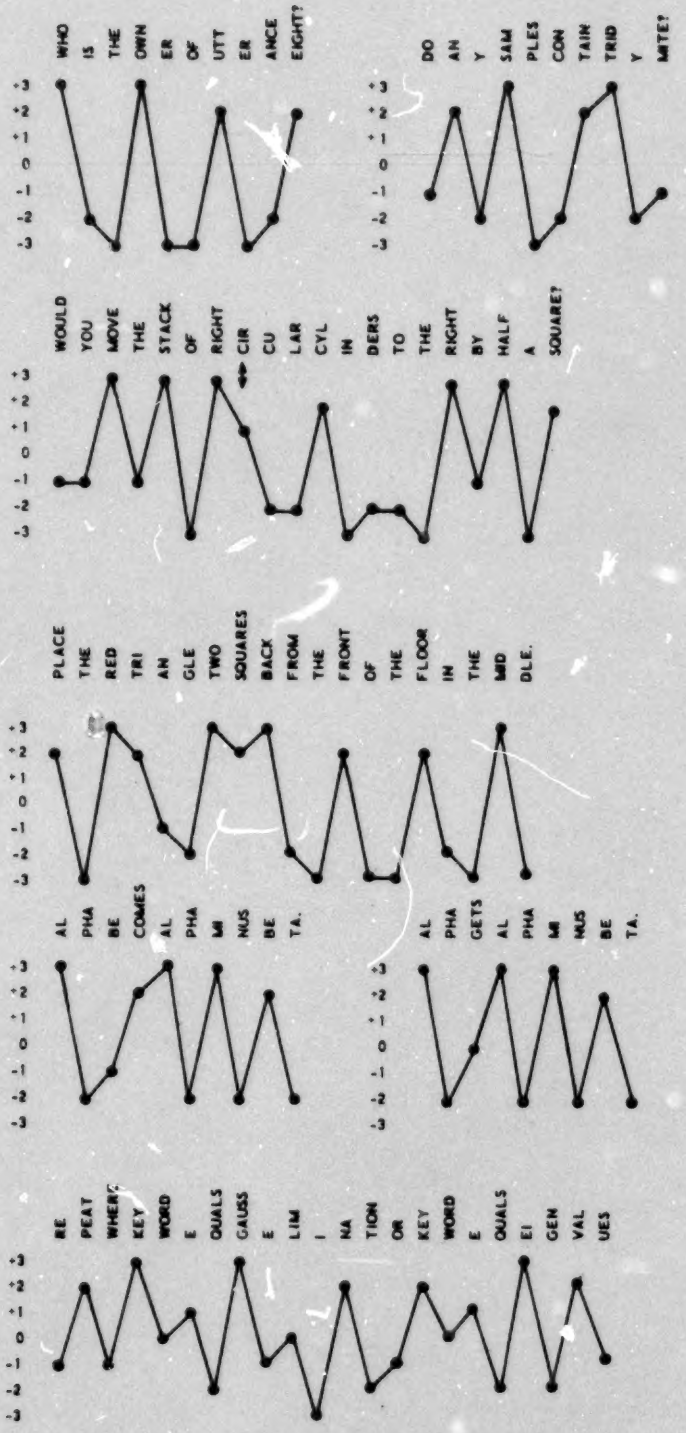

WHO IS THE OWN ER OF UTT ER ANCE EIGHT?
DO AN Y SAM PLES CON TAIN TRID Y MTE?
WOULD YOU MOVE THE STACK OF RIGHT CIR CU LAR CYL IN DERS TO THE RIGHT BY HALF A SQUARE?
PLACE THE RED TRI AN GLE TWO SQUARES BACK FROM THE FRONT OF THE FLOOR IN THE MID DLE.
AL PHA BE COMES AL PHA MI NUS BE TA.
AL PHA GETS AL PHA MI NUS BE TA.
RE PEAT WHERE KEY WORD E QUALS GAUSS E LIM I NA TION OR KEY WORD E QUALS EI GEN VAL UES

Figure B-14. Summary of Stress Judgments by Three Listeners, for the 7ARPA Sentences as Spoken.

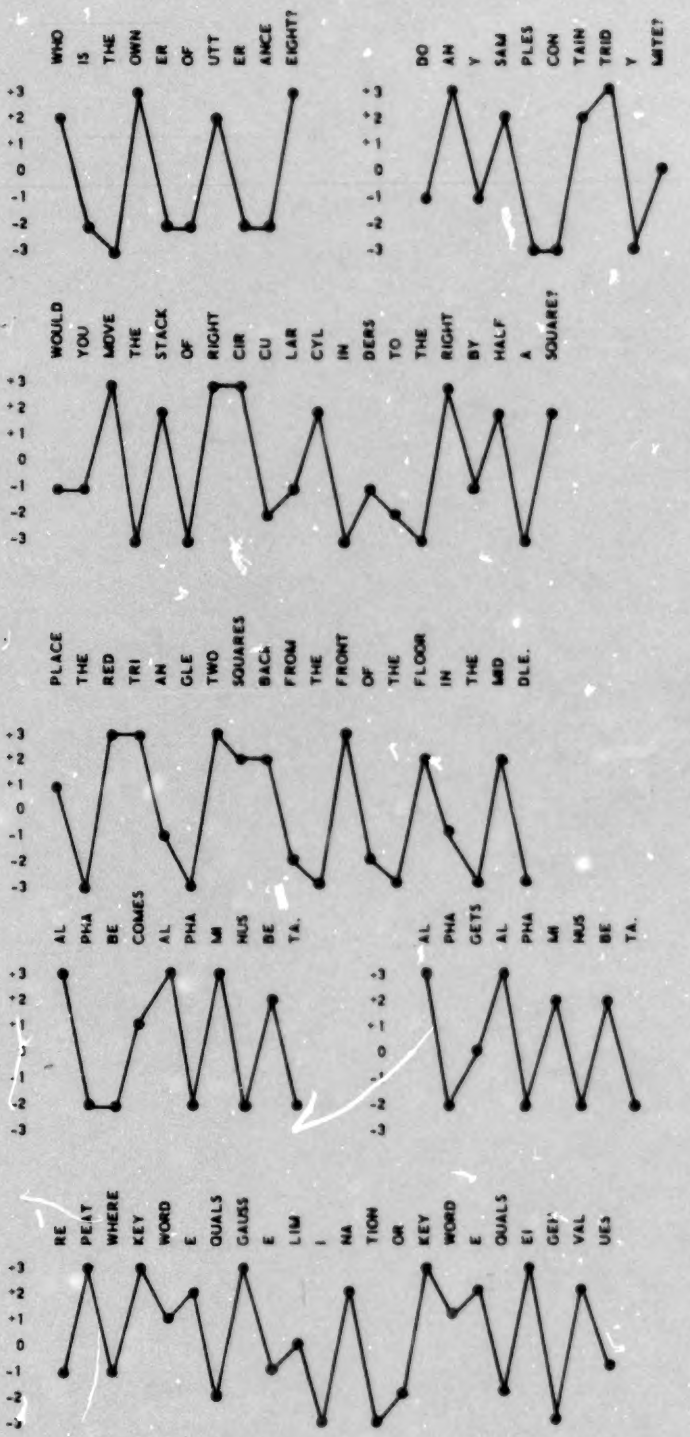
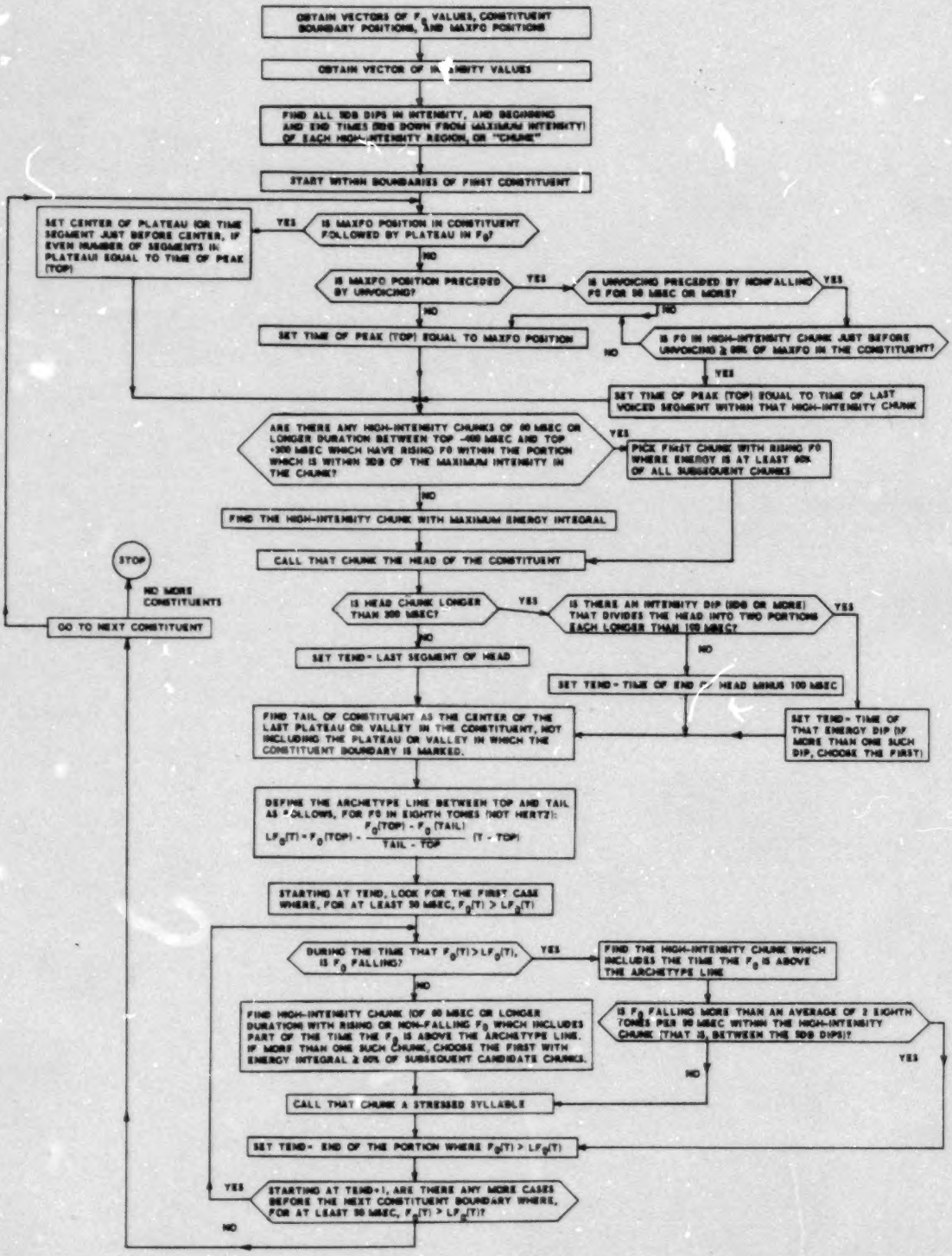

WHO IS THE OWN ER OF UTT ER ANCE EIGHT?
DO AN Y SAM PLES CON TAIN TRID Y MTE?
WOULD YOU MOVE THE STACK OF RIGHT CIR CU LAR CYL IN DERS TO THE RIGHT BY HALF A SQUARE?
PLACE THE RED TRI AN GLE TWO SQUARES BACK FROM THE FRONT OF THE FLOOR IN THE MID DLE.
AL PHA BE COMES AL PHA MI NUS BE TA.
AL PHA GETS AL PHA MI NUS BE TA.
RE PEAT WHERE KEY WORD E QUALS GAUSS E LIM I NA TION OR KEY WORD E QUALS EI GE' VAL UES

Figure B-15. Summary of Stress Judgments by Three 'Listeners', When Given Only the Written Text of the 7ARPA Sentences (NO SPEECH).

APPENDIX C: STRESSED SYLLABLES LOCATED BY ALGORITHM

A flowchart of the stressed syllable location algorithm is shown in Figure C-1. This is a characterization of the hand analysis procedure, and may have to be modified and specified in more detail for implementation as a computer program.

The results of applying the algorithm to stressed syllable location for each of the recorded speech texts are shown in Figures C-2 to C-11. The figures show the majority stress scores above each syllable. Those syllables perceived as stressed by two or more listeners (i.e., SS = +2 or +3) are shown in boxes. The syllables or speech portions which were declared to be stressed by the algorithm are shown underlined. Whenever an underlined portion includes a boxed-in stressed syllable, a correct location has been obtained. Cases where an underlined portion did not include a boxed-in syllable (that is, no part was perceived as stressed by two or more listeners) are false locations of stressed syllables. Many of these false locations resulted from false constituent boundary detections, since the present procedure demands that every detected constituent have a stressed HEAD.


OBTAIN VECTORS OF F_0 VALUES, CONSTITUENT BOUNDARY POSITIONS, AND MAXFO POSITIONS
OBTAIN VECTOR OF INTENSITY VALUES
FIND ALL 10DB DIPS IN INTENSITY, AND BEGINNING AND END TIMES (DB DOWN FROM MAXIMUM INTENSITY) OF EACH HIGH-INTENSITY REGION, OR "CHUNK"
START WITHIN BOUNDARIES OF FIRST CONSTITUENT
IS MAXFO POSITION IN CONSTITUENT FOLLOWED BY PLATEAU IN F_0?
YES
SET CENTER OF PLATEAU (OR TIME SEGMENT JUST BEFORE CENTER, IF EVEN NUMBER OF SEGMENTS IN PLATEAU) EQUAL TO TIME OF PEAK (TOP)
NO
IS MAXFO POSITION PRECEDED BY UNVOICING?
YES
IS UNVOICING PRECEDED BY NONFALLING F_0 FOR 30 MSEC OR MORE?
YES
NO
NO
SET TIME OF PEAK (TOP) EQUAL TO MAXFO POSITION
IS F_0 IN HIGH-INTENSITY CHUNK JUST BEFORE UNVOICING ≥ 88% OF MAXFO IN THE CONSTITUENT?
NO
YES
SET TIME OF PEAK (TOP) EQUAL TO TIME OF LAST VOICED SEGMENT WITHIN THAT HIGH-INTENSITY CHUNK
ARE THERE ANY HIGH-INTENSITY CHUNKS OF 60 MSEC OR LONGER DURATION BETWEEN TOP -400 MSEC AND TOP +300 MSEC WHICH HAVE RISING F_0 WITHIN THE PORTION WHICH IS WITHIN 3DB OF THE MAXIMUM INTENSITY IN THE CHUNK?
YES
PICK FIRST CHUNK WITH RISING F_0 WHERE ENERGY IS AT LEAST 60% OF ALL SUBSEQUENT CHUNKS
NO
FIND THE HIGH-INTENSITY CHUNK WITH MAXIMUM ENERGY INTEGRAL
CALL THAT CHUNK THE HEAD OF THE CONSTITUENT
IS HEAD CHUNK LONGER THAN 300 MSEC?
YES
IS THERE AN INTENSITY DIP (3DB OR MORE) THAT DIVIDES THE HEAD INTO TWO PORTIONS EACH LONGER THAN 100 MSEC?
YES
NO
SET TEND = LAST SEGMENT OF HEAD
NO
SET TEND = TIME OF END OF HEAD MINUS 100 MSEC
SET TEND = TIME OF THAT ENERGY DIP (IF MORE THAN ONE SUCH DIP, CHOOSE THE FIRST)
FIND TAIL OF CONSTITUENT AS THE CENTER OF THE LAST PLATEAU OR VALLEY IN THE CONSTITUENT, NOT INCLUDING THE PLATEAU OR VALLEY IN WHICH THE CONSTITUENT BOUNDARY IS MARKED.
DEFINE THE ARCHETYPE LINE BETWEEN TOP AND TAIL AS FOLLOWS, FOR F0 IN EIGHTH TONES (NOT HERTZ):
$LF_0(T) = F_0(TOP) - \frac{F_0(TOP) - F_0(TAIL)}{TAIL - TOP} (T - TOP)$
STARTING AT TEND, LOOK FOR THE FIRST CASE WHERE, FOR AT LEAST 30 MSEC, $F_0(T) > LF_0(T)$
DURING THE TIME THAT $F_0(T) > LF_0(T)$, IS F_0 FALLING?
YES
FIND THE HIGH-INTENSITY CHUNK WHICH INCLUDES THE TIME THE F_0 IS ABOVE THE ARCHETYPE LINE
NO
FIND HIGH-INTENSITY CHUNK (OF 60 MSEC OR LONGER DURATION) WITH RISING OR NON-FALLING F_0 WHICH INCLUDES PART OF THE TIME THE F_0 IS ABOVE THE ARCHETYPE LINE. IF MORE THAN ONE SUCH CHUNK, CHOOSE THE FIRST WITH ENERGY INTEGRAL ≥ 60% OF SUBSEQUENT CANDIDATE CHUNKS.
IS F_0 FALLING MORE THAN AN AVERAGE OF 2 EIGHTH TONES PER 90 MSEC WITHIN THE HIGH-INTENSITY CHUNK (THAT IS, BETWEEN THE 3DB DIPS)?
YES
NO
CALL THAT CHUNK A STRESSED SYLLABLE
SET TEND = END OF THE PORTION WHERE $F_0(T) > LF_0(T)$
STARTING AT TEND+1, ARE THERE ANY MORE CASES BEFORE THE NEXT CONSTITUENT BOUNDARY WHERE, FOR AT LEAST 30 MSEC, $F_0(T) > LF_0(T)$?
YES
NO
NO MORE CONSTITUENTS
GO TO NEXT CONSTITUENT
STOP

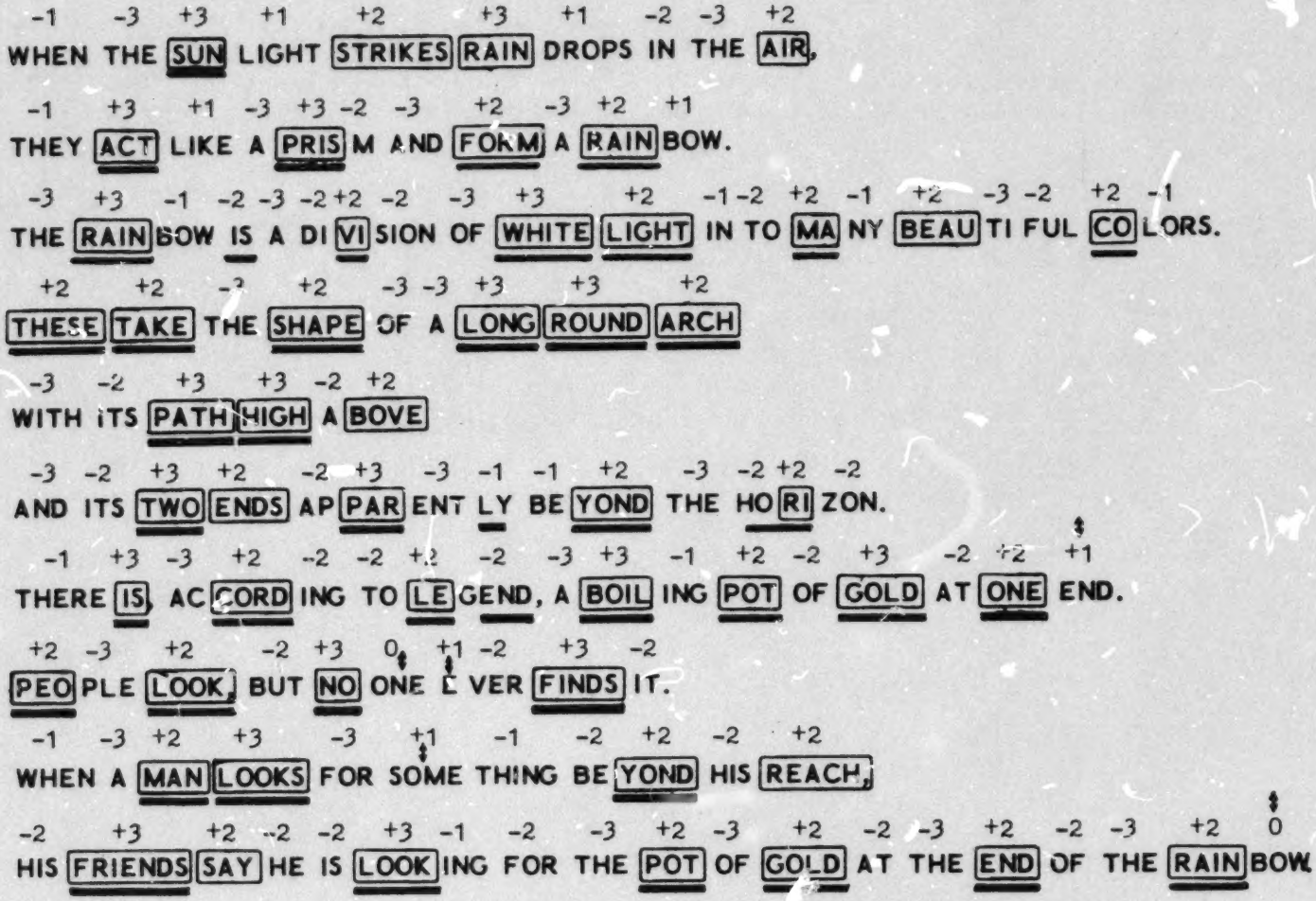

-1 -3 +3 +1 +2 +3 +1 -2 -3 +2
WHEN THE SUN LIGHT STRIKES RAIN DROPS IN THE AIR,
-1 +3 +1 -3 +3 -2 -3 +2 -3 +2 +1
THEY ACT LIKE A PRIS M AND FORM A RAIN BOW.
-3 +3 -1 -2 -3 -2 +2 -2 -3 +3 +2 -1 -2 +2 -1 +2 -3 -2 +2 -1
THE RAIN BOW IS A DI VI SION OF WHITE LIGHT IN TO MA NY BEAU TI FUL CO LORS.
+2 +2 -3 +2 -3 -3 +3 +3 +2
THESE TAKE THE SHAPE OF A LONG ROUND ARCH
-3 -2 +3 +3 -2 +2
WITH ITS PATH HIGH A BOVE
-3 -2 +3 +2 -2 +3 -3 -1 -1 +2 -3 -2 +2 -2
AND ITS TWO ENDS AP PAR ENT LY BE YOND THE HO RI ZON.
-1 +3 -3 +2 -2 -2 +2 -2 -3 +3 -1 +2 -2 +3 -2 +2 +1
THERE IS AC CORD ING TO LE GEND, A BOIL ING POT OF GOLD AT ONE END.
+2 -3 +2 -2 +3 0 +1 -2 +3 -2
PEO PLE LOOK BUT NO ONE E VER FINDS IT.
-1 -3 +2 +3 -3 +1 -1 -2 +2 -2 +2
WHEN A MAN LOOKS FOR SOME THING BE YOND HIS REACH,
-2 +3 +2 -2 -2 +3 -1 -2 -3 +2 -3 +2 -2 -3 +2 -2 -3 +2 0
HIS FRIENDS SAY HE IS LOOK ING FOR THE POT OF GOLD AT THE END OF THE RAIN BOW.

Figure C-2. Comparison of Algorithmically Located Stressed "Syllables" with Perceived Stress Patterns, for Talker ASH Reading the Rainbow Script. Stress scores (SS) are shown above each Syllable. Syllables perceived as stressed by two or more listeners (SS= + 2 or + 3) are enclosed in boxes. Portions located by the algorithm are underlined.

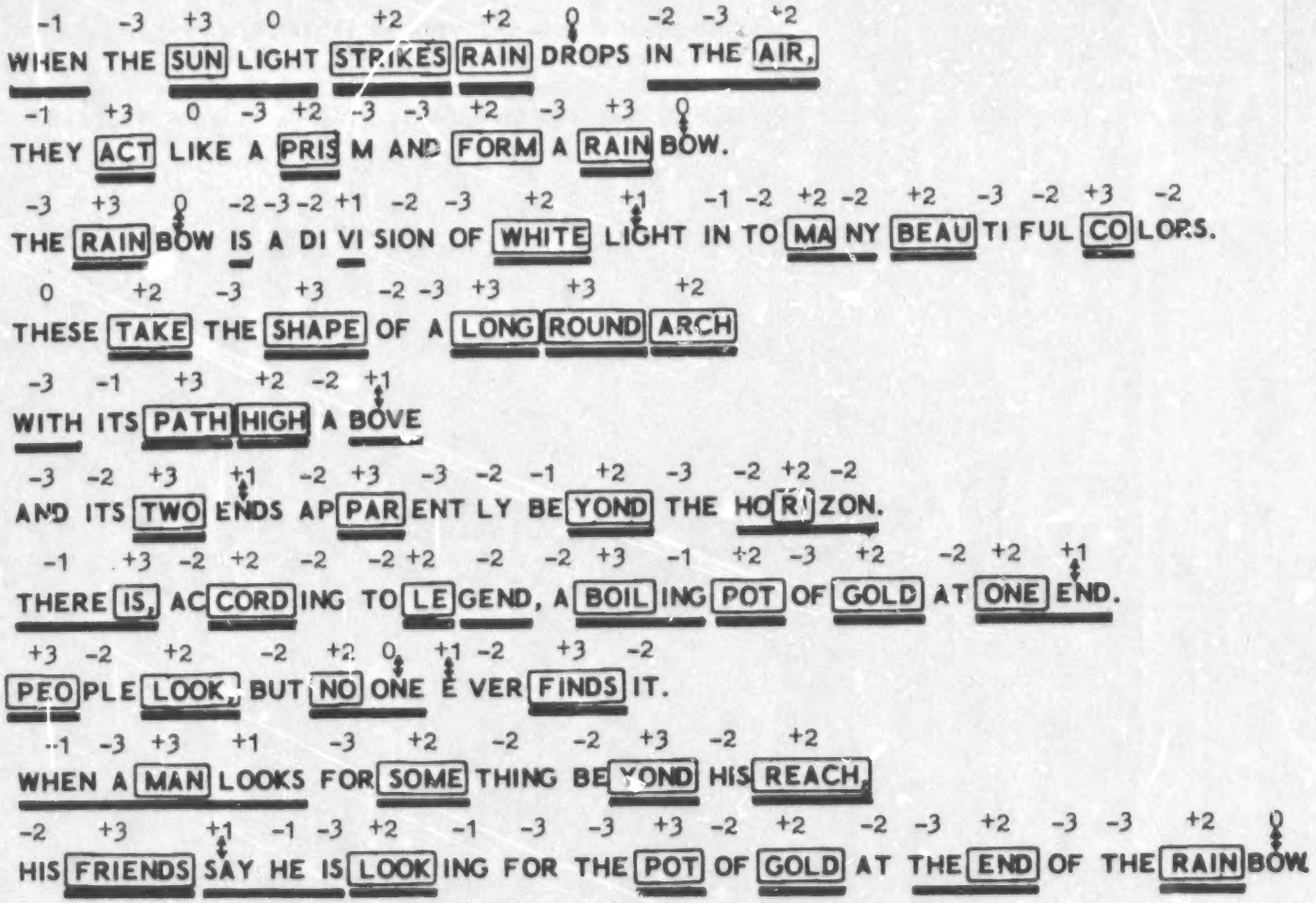

-1 -3 +3 0 +2 +2 0 -2 -3 +2
WHEN THE SUN LIGHT STRIKES RAIN DROPS IN THE AIR,
-1 +3 0 -3 +2 -3 -3 +2 -3 +3 0
THEY ACT LIKE A PRIS M AND FORM A RAIN BOW.
-3 +3 0 -2 -3 -2 +1 -2 -3 +2 +1 -1 -2 +2 -2 +2 -3 -2 +3 -2
THE RAIN BOW IS A DI VI SION OF WHITE LIGHT IN TO MA NY BEAU TI FUL CO LORS.
0 +2 -3 +3 -2 -3 +3 +3 +2
THESE TAKE THE SHAPE OF A LONG ROUND ARCH
-3 -1 +3 +2 -2 +1
WITH ITS PATH HIGH A BOVE
-3 -2 +3 +1 -2 +3 -3 -2 -1 +2 -3 -2 +2 -2
AND ITS TWO ENDS AP PAR ENT LY BE YOND THE HO RI ZON.
-1 +3 -2 +2 -2 -2 +2 -2 -2 +3 -1 +2 -3 +2 -2 +2 +1
THERE IS, AC CORD ING TO LE GEND, A BOIL ING POT OF GOLD AT ONE END.
+3 -2 +2 -2 +2 0 +1 -2 +3 -2
PEO PLE LOOK, BUT NO ONE E VER FINDS IT.
-1 -3 +3 +1 -3 +2 -2 -2 +3 -2 +2
WHEN A MAN LOOKS FOR SOME THING BE YOND HIS REACH,
-2 +3 +1 -1 -3 +2 -1 -3 -3 +3 -2 +2 -2 -3 +2 -3 -3 +2 0
HIS FRIENDS SAY HE IS LOOK ING FOR THE POT OF GOLD AT THE END OF THE RAIN BOW.

Figure C-3. Comparison of Algorithmically Located Stressed "Syllables" with Perceived Stress Patterns, for Talker GWH Reading the Rainbow Script. Stress scores (SS) are shown above each Syllable. Syllables perceived as stressed by two or more listeners (SS= + 2 or + 3) are enclosed in boxes. Portions located by the algorithm are underlined.

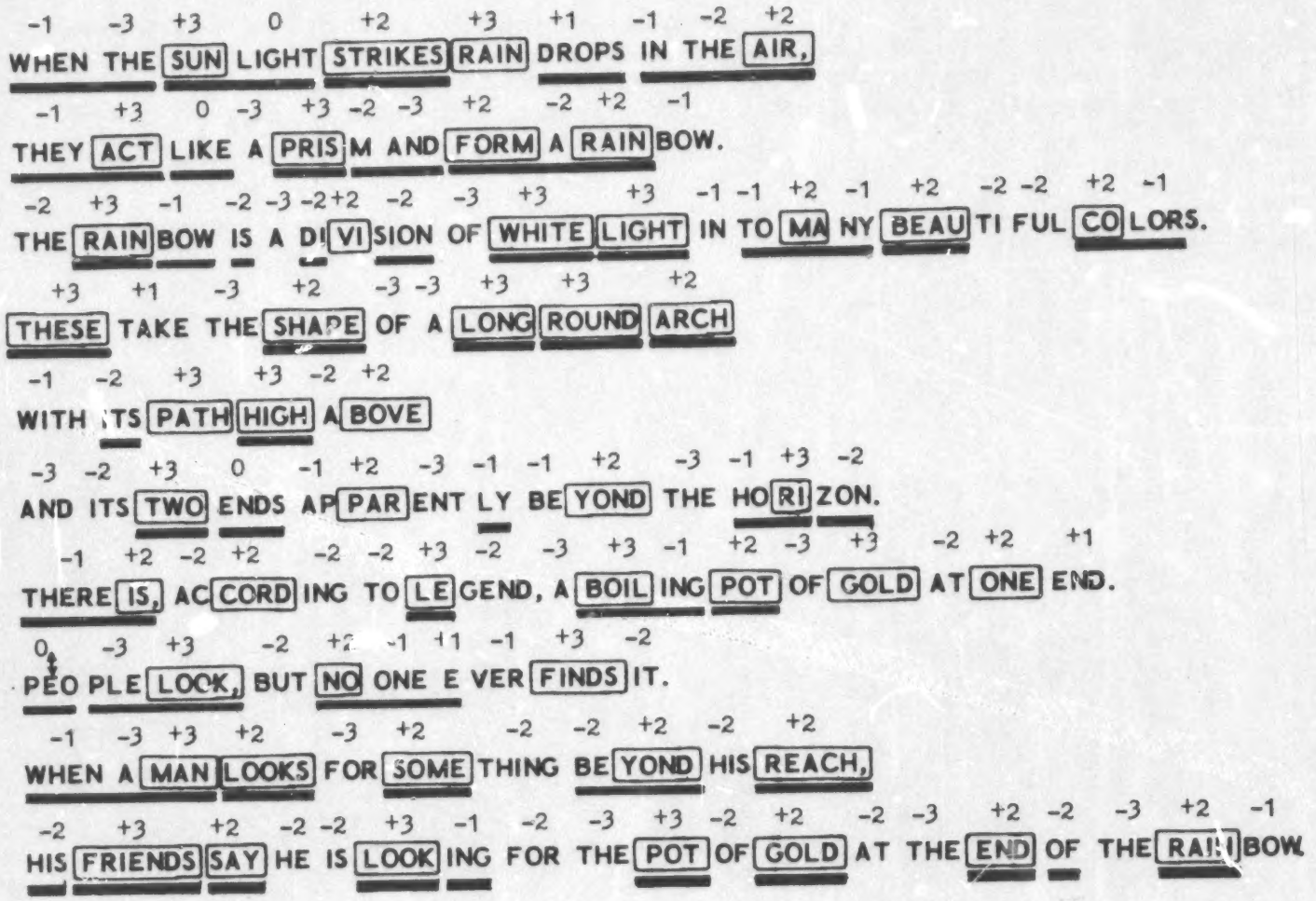

-1 -3 +3 0 +2 +3 +1 -1 -2 +2
WHEN THE SUN LIGHT STRIKES RAIN DROPS IN THE AIR,
-1 +3 0 -3 +3 -2 -3 +2 -2 +2 -1
THEY ACT LIKE A PRIS M AND FORM A RAIN BOW.
-2 +3 -1 -2 -3 -2 +2 -2 -3 +3 +3 -1 -1 +2 -1 +2 -2 -2 +2 -1
THE RAIN BOW IS A DI VI SION OF WHITE LIGHT IN TO MA NY BEAU TI FUL CO LORS.
+3 +1 -3 +2 -3 -3 +3 +3 +2
THESE TAKE THE SHAPE OF A LONG ROUND ARCH
-1 -2 +3 +3 -2 +2
WITH ITS PATH HIGH A BOVE
-3 -2 +3 0 -1 +2 -3 -1 -1 +2 -3 -1 +3 -2
AND ITS TWO ENDS AP PAR ENT LY BE YOND THE HO RI ZON.
-1 +2 -2 +2 -2 -2 +3 -2 -3 +3 -1 +2 -3 +3 -2 +2 +1
THERE IS, AC CORD ING TO LE GEND, A BOIL ING POT OF GOLD AT ONE END.
0 -3 +3 -2 +2 -1 +1 -1 +3 -2
PEO PLE LOOK, BUT NO ONE E VER FINDS IT.
-1 -3 +3 +2 -3 +2 -2 -2 +2 -2 +2
WHEN A MAN LOOKS FOR SOME THING BE YOND HIS REACH,
-2 +3 +2 -2 -2 +3 -1 -2 -3 +3 -2 +2 -2 -3 +2 -2 -3 +2 -1
HIS FRIENDS SAY HE IS LOOK ING FOR THE POT OF GOLD AT THE END OF THE RAINBOW.

Figure C-4. Comparison of Algorithmically Located Stressed "Syllables" with Perceived Stress Patterns, for Talker WB Reading the Rainbow Script. Stress scores (SS) are shown above each Syllable. Syllables perceived as stressed by two or more listeners (SS= + 2 or + 3) are enclosed in boxes. Portions located by the algorithm are underlined.

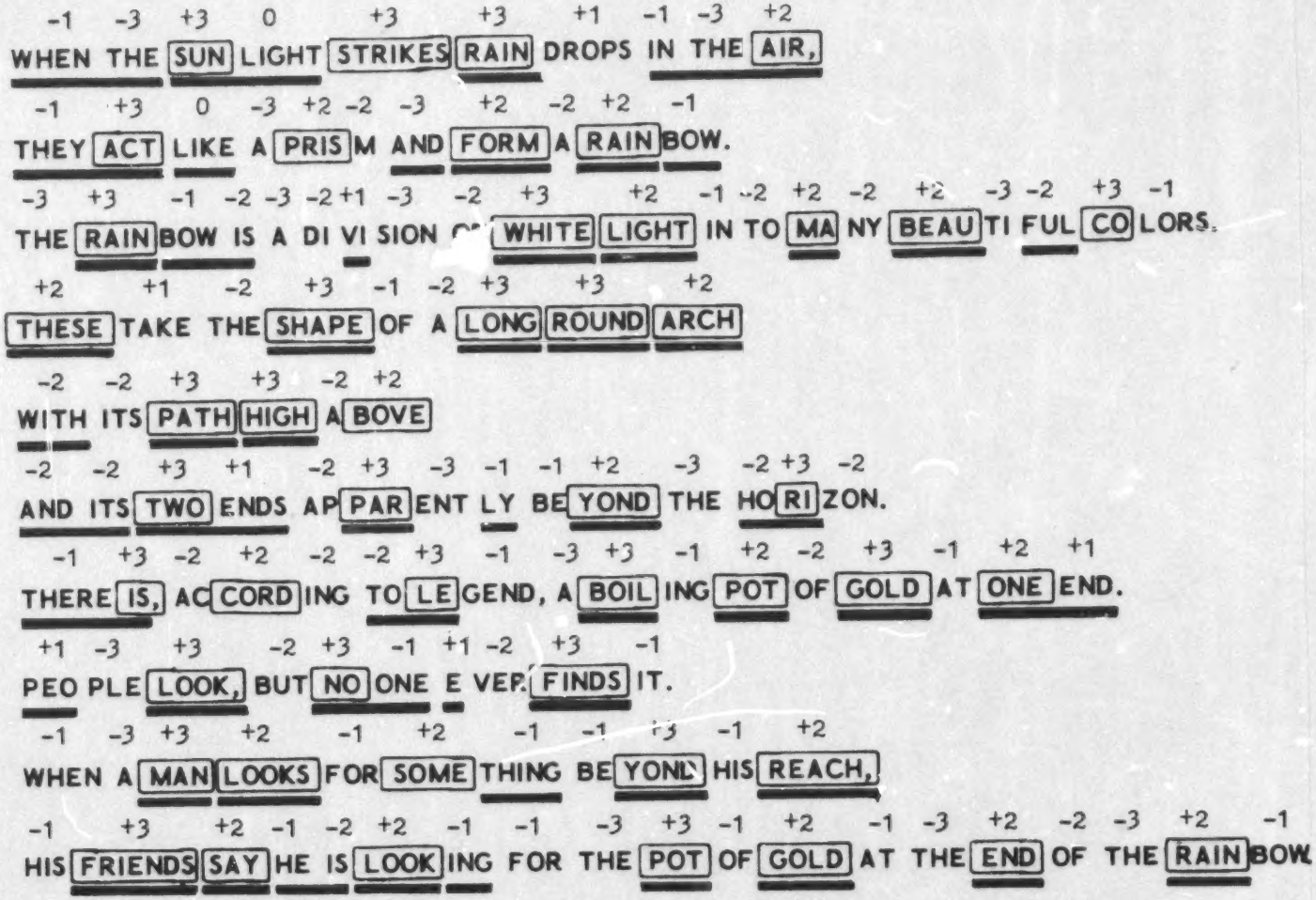

-1 -3 +3 0 +3 +3 +1 -1 -3 +2
WHEN THE SUN LIGHT STRIKES RAIN DROPS IN THE AIR,
-1 +3 0 -3 +2 -2 -3 +2 -2 +2 -1
THEY ACT LIKE A PRIS M AND FORM A RAIN BOW.
-3 +3 -1 -2 -3 -2 +1 -3 -2 +3 +2 -1 -2 +2 -2 +2 -3 -2 +3 -1
THE RAIN BOW IS A DI VI SION OF WHITE LIGHT IN TO MA NY BEAU TI FUL CO LORS.
+2 +1 -2 +3 -1 -2 +3 +3 +2
THESE TAKE THE SHAPE OF A LONG ROUND ARCH
-2 -2 +3 +3 -2 +2
WITH ITS PATH HIGH A BOVE
-2 -2 +3 +1 -2 +3 -3 -1 -1 +2 -3 -2 +3 -2
AND ITS TWO ENDS AP PAR ENT LY BE YOND THE HO RI ZON.
-1 +3 -2 +2 -2 -2 +3 -1 -3 +3 -1 +2 -2 +3 -1 +2 +1
THERE IS, AC CORD ING TO LE GEND, A BOIL ING POT OF GOLD AT ONE END.
+1 -3 +3 -2 +3 -1 +1 -2 +3 -1
PEO PLE LOOK, BUT NO ONE E VER FINDS IT.
-1 -3 +3 +2 -1 +2 -1 -1 +3 -1 +2
WHEN A MAN LOOKS FOR SOME THING BE YOND HIS REACH,
-1 +3 +2 -1 -2 +2 -1 -1 -3 +3 -1 +2 -1 -3 +2 -2 -3 +2 -1
HIS FRIENDS SAY HE IS LOOK ING FOR THE POT OF GOLD AT THE END OF THE RAIN BOW.

Figure C-5. Comparison of Algorithmically Located Stressed "Syllables" with Perceived Stress Patterns, for Talker JP Reading the Rainbow Script. Stress scores (SS) are shown above each Syllable. Syllables perceived as stressed by two or more listeners (SS= + 2 or + 3) are enclosed in boxes. Portions located by the algorithm are underlined.

-1 -3 +3 0 +2 +3 0 -2 -3 +2
WHEN THE [SUN] LIGHT [STRIKES] [RAIN] DROPS IN THE [AIR,]
-1 +2 0 -3 +3 -2 -3 +2 -2 +3 0
THEY [ACT] LIKE A [PRIS]M AND [FORM] A [RAIN]BOW.
-2 +3 0 -2 -3 -2 +1 -2 -2 +3 +3 -1 -2 +3 -1 +2 -3 -2 +2 -1
THE [RAIN]BOW IS A DI [VI] SION OF [WHITE] [LIGHT] IN TO [MA]NY [BEAU]TI FUL [CO]LORS.
+1 +1 -3 +2 -3 -2 +3 +3 +2
THESE TAKE THE [SHAPE] OF A [LONG] [ROUND] [ARCH]
-2 -2 +3 +3 -2 +2
WITH ITS [PATH] [HIGH] A[BOVE]
-2 -1 +3 +1 -2 +3 -3 -2 -2 +2 -3 -1 +3 -2
AND ITS [TWO] ENDS AP[PAR]ENT LY BE[YOND] THE HO[RI]ZON.
-1 +3 -2 +3 -1 -3 +3 -2 -3 +3 -1 +2 -2 +2 -2 +3 +2
THERE [IS,] AC[CORD]ING TO [LE]GEND, A [BOIL]ING [POT] OF [GOLD] AT [ONE] [END.]
+2 -2 +2 -2 +3 0 +2 -1 +3 -2
[PEO]PLE [LOOK,] BUT [NO] ONE [E]VER [FINDS] IT.
-1 -3 +3 +2 -3 +2 -1 -2 +3 -1 +2
WHEN A [MAN] [LOOKS] FOR [SOME] THING BE[YOND] HIS [REACH,]
-2 +3 +2 -2 -3 +2 -2 -2 -3 +3 -3 +2 -2 -2 +3 -3 -2 +3 0
HIS [FRIENDS] [SAY] HE IS [LOOK]ING FOR THE [POT] OF [GOLD] AT THE [END] OF THE [RAIN]BOW.

Figure C-6. Comparison of Algorithmically Located Stressed "Syllables" with Perceived Stress Patterns, for Talker PB Reading the Rainbow Script. Stress scores (SS) are shown above each Syllable. Syllables perceived as stressed by two or more listeners (SS = + 2 or + 3) are enclosed in boxes. Portions located by the algorithm are underlined.

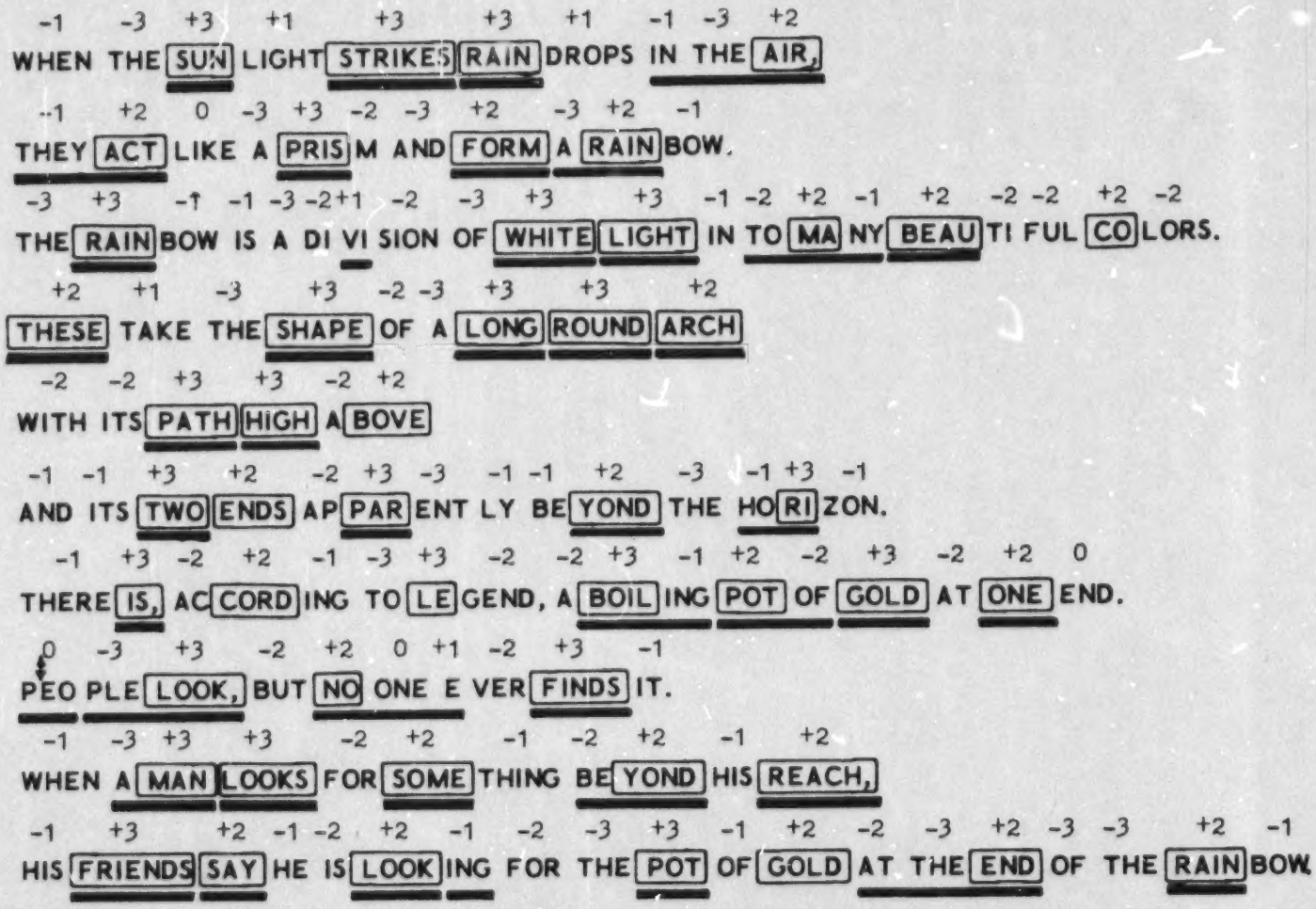

-1 -3 +3 +1 +3 +3 +1 -1 -3 +2
WHEN THE SUN LIGHT STRIKES RAIN DROPS IN THE AIR,
-1 +2 0 -3 +3 -2 -3 +2 -3 +2 -1
THEY ACT LIKE A PRIS M AND FORM A RAIN BOW.
-3 +3 -1 -1 -3 -2 +1 -2 -3 +3 +3 -1 -2 +2 -1 +2 -2 -2 +2 -2
THE RAIN BOW IS A DI VI SION OF WHITE LIGHT IN TO MA NY BEAU TI FUL CO LORS.
+2 +1 -3 +3 -2 -3 +3 +3 +2
THESE TAKE THE SHAPE OF A LONG ROUND ARCH
-2 -2 +3 +3 -2 +2
WITH ITS PATH HIGH A BOVE
-1 -1 +3 +2 -2 +3 -3 -1 -1 +2 -3 -1 +3 -1
AND ITS TWO ENDS AP PAR ENT LY BE YOND THE HO RI ZON.
-1 +3 -2 +2 -1 -3 +3 -2 -2 +3 -1 +2 -2 +3 -2 +2 0
THERE IS, AC CORD ING TO LE GEND, A BOIL ING POT OF GOLD AT ONE END.
0 -3 +3 -2 +2 0 +1 -2 +3 -1
PEO PLE LOOK, BUT NO ONE E VER FINDS IT.
-1 -3 +3 +3 -2 +2 -1 -2 +2 -1 +2
WHEN A MAN LOOKS FOR SOME THING BE YOND HIS REACH,
-1 +3 +2 -1 -2 +2 -1 -2 -3 +3 -1 +2 -2 -3 +2 -3 -3 +2 -1
HIS FRIENDS SAY HE IS LOOK ING FOR THE POT OF GOLD AT THE END OF THE RAIN BOW.

Figure C-7. Comparison of Algorithmically Located Stressed "Syllables" with Perceived Stress Patterns, for Talker ER Reading the Rainbow Script. Stress scores (SS) are shown above each Syllable. Syllables perceived as stressed by two or more listeners (SS= + 2 or + 3) are enclosed in boxes. Portions located by the algorithm are underlined.

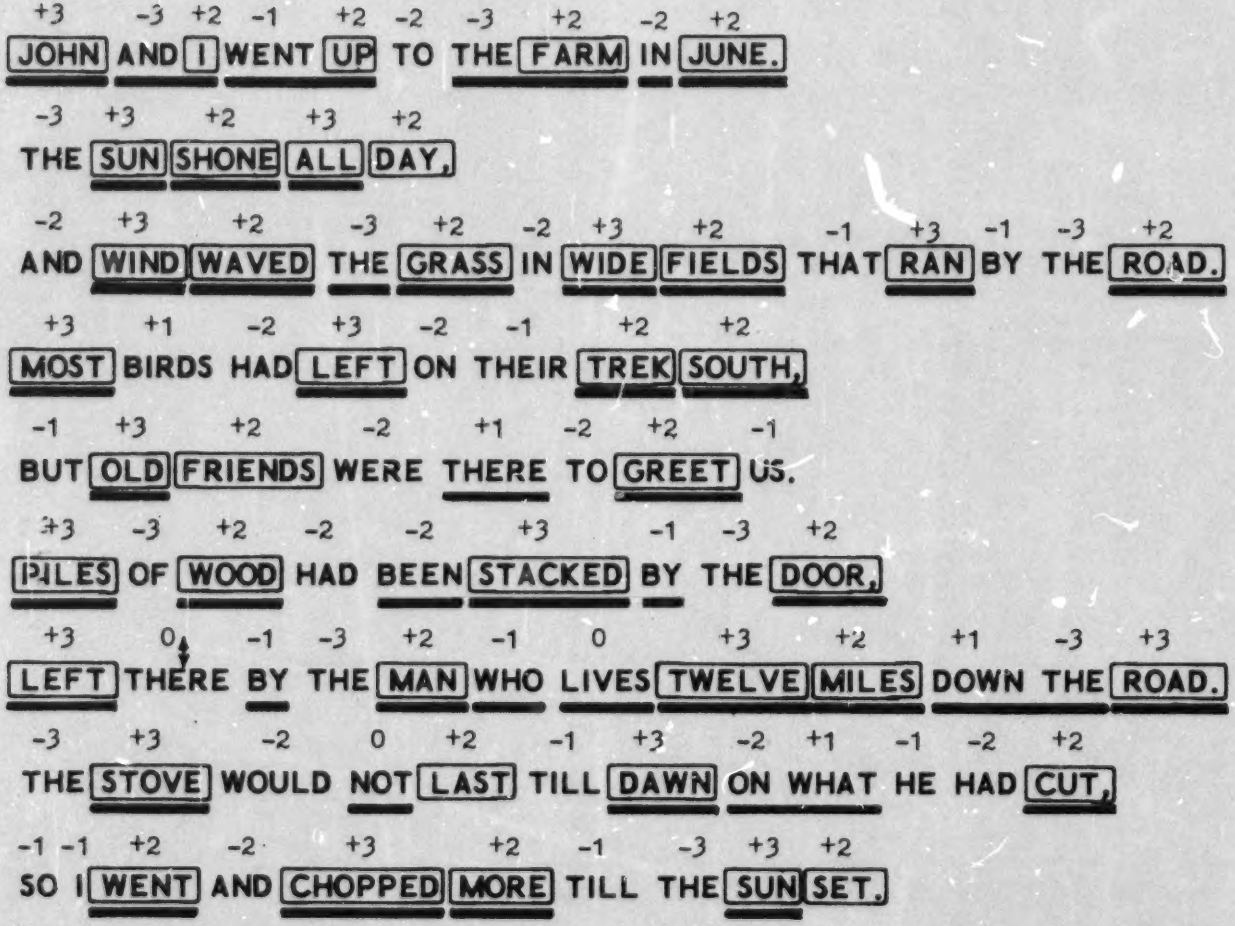

+3 -3 +2 -1 +2 -2 -3 +2 -2 +2
JOHN AND I WENT UP TO THE FARM IN JUNE.
-3 +3 +2 +3 +2
THE SUN SHONE ALL DAY,
-2 +3 +2 -3 +2 -2 +3 +2 -1 +3 -1 -3 +2
AND WIND WAVED THE GRASS IN WIDE FIELDS THAT RAN BY THE ROAD.
+3 +1 -2 +3 -2 -1 +2 +2
MOST BIRDS HAD LEFT ON THEIR TREK SOUTH,
-1 +3 +2 -2 +1 -2 +2 -1
BUT OLD FRIENDS WERE THERE TO GREET US.
+3 -3 +2 -2 -2 +3 -1 -3 +2
PILES OF WOOD HAD BEEN STACKED BY THE DOOR,
+3 0 -1 -3 +2 -1 0 +3 +2 +1 -3 +3
LEFT THERE BY THE MAN WHO LIVES TWELVE MILES DOWN THE ROAD.
-3 +3 -2 0 +2 -1 +3 -2 +1 -1 -2 +2
THE STOVE WOULD NOT LAST TILL DAWN ON WHAT HE HAD CUT,
-1 -1 +2 -2 +3 +2 -1 -3 +3 +2
SO I WENT AND CHOPPED MORE TILL THE SUN SET.

Figure C-8. Comparison of Algorithmically Located Stressed "Syllables" with Perceived Stress Patterns, for Talker ASH Reading the Monosyllabic Script.

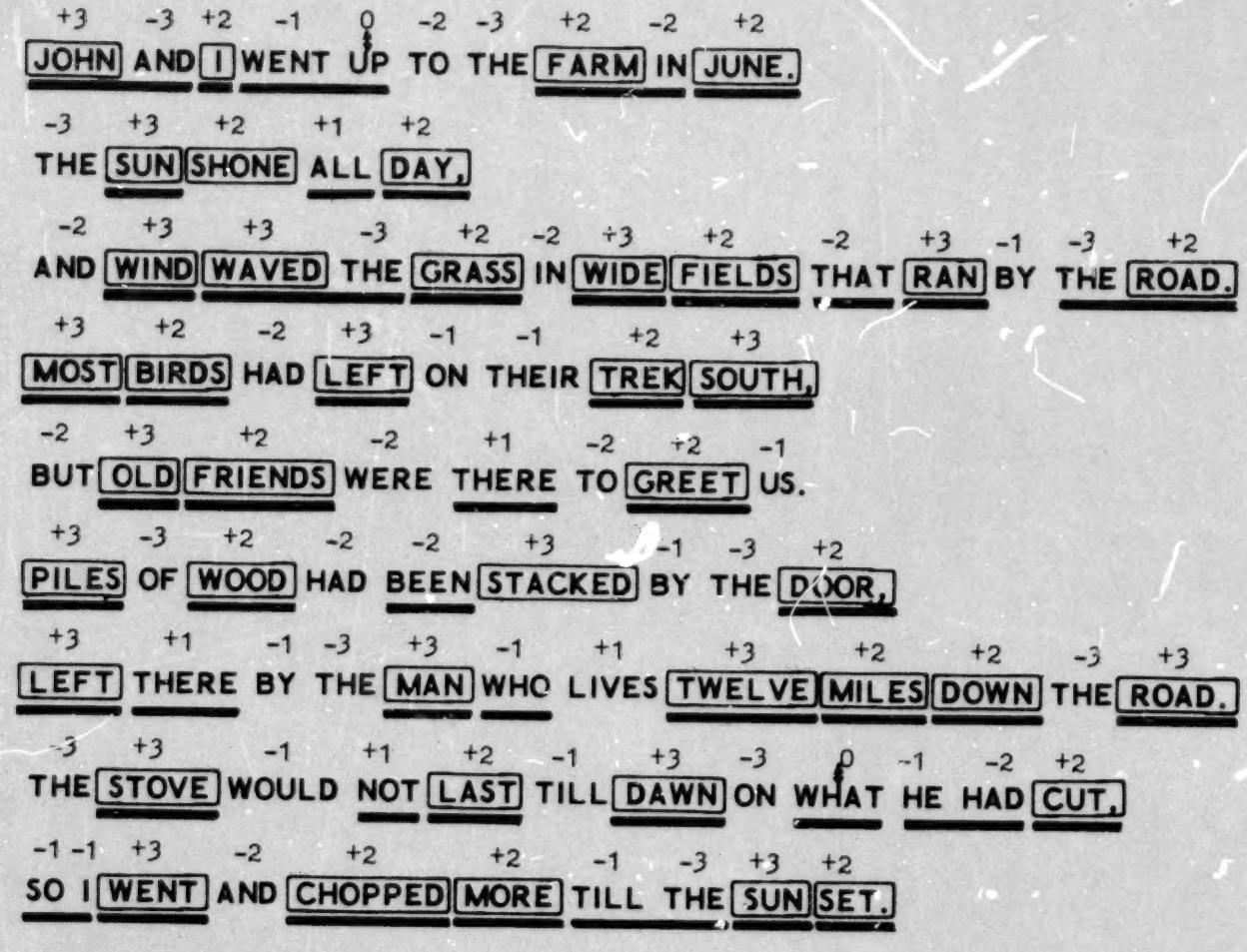

+3 -3 +2 -1 0 -2 -3 +2 -2 +2
JOHN AND I WENT UP TO THE FARM IN JUNE.
-3 +3 +2 +1 +2
THE SUN SHONE ALL DAY,
-2 +3 +3 -3 +2 -2 +3 +2 -2 +3 -1 -3 +2
AND WIND WAVED THE GRASS IN WIDE FIELDS THAT RAN BY THE ROAD.
+3 +2 -2 +3 -1 -1 +2 +3
MOST BIRDS HAD LEFT ON THEIR TREK SOUTH,
-2 +3 +2 -2 +1 -2 +2 -1
BUT OLD FRIENDS WERE THERE TO GREET US.
+3 -3 +2 -2 -2 +3 -1 -3 +2
PILES OF WOOD HAD BEEN STACKED BY THE DOOR,
+3 +1 -1 -3 +3 -1 +1 +3 +2 +2 -3 +3
LEFT THERE BY THE MAN WHO LIVES TWELVE MILES DOWN THE ROAD.
-3 +3 -1 +1 +2 -1 +3 -3 0 -1 -2 +2
THE STOVE WOULD NOT LAST TILL DAWN ON WHAT HE HAD CUT,
-1 -1 +3 -2 +2 +2 -1 -3 +3 +2
SO I WENT AND CHOPPED MORE TILL THE SUN SET.

Figure C-9. Comparison of Algorithmically Located Stressed "Syllables" with Perceived Stress Patterns, for Talker GWH Reading the Monosyllabic Script.

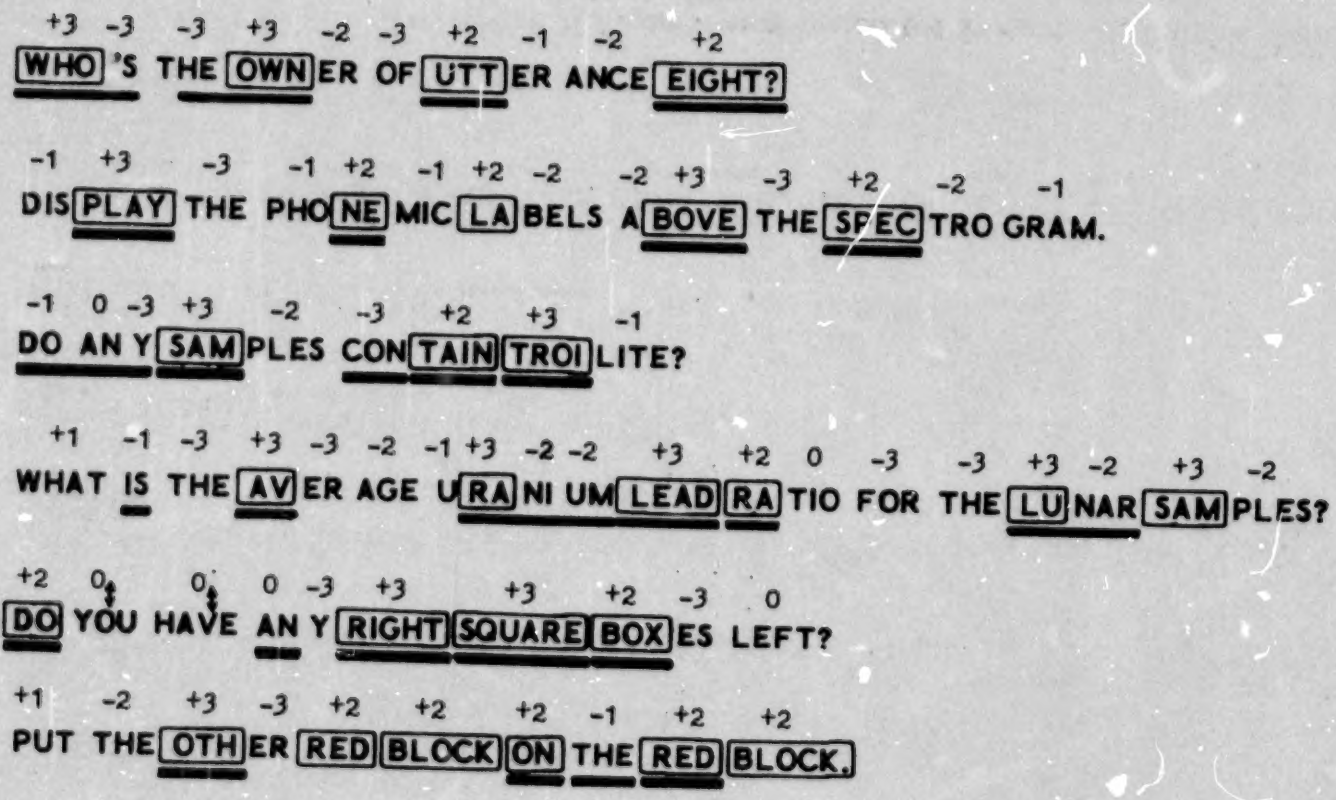

+3 -3 -3 +3 -2 -3 +2 -1 -2 +2
WHO 'S THE OWN ER OF UTT ER ANCE EIGHT?
-1 +3 -3 -1 +2 -1 +2 -2 -2 +3 -3 +2 -2 -1
DIS PLAY THE PHO NE MIC LA BELS A BOVE THE SPEC TRO GRAM.
-1 0 -3 +3 -2 -3 +2 +3 -1
DO AN Y SAM PLES CON TAIN TROI LITE?
+1 -1 -3 +3 -3 -2 -1 +3 -2 -2 +3 +2 0 -3 -3 +3 -2 +3 -2
WHAT IS THE AV ER AGE U RA NI UM LEAD RA TIO FOR THE LU NAR SAM PLES?
+2 0 0 0 -3 +3 +3 +2 -3 0
DO YOU HAVE AN Y RIGHT SQUARE BOX ES LEFT?
+1 -2 +3 -3 +2 +2 +2 -1 +2 +2
PUT THE OTH ER RED BLOCK ON THE RED BLOCK.

Figure C-10. Comparison of Algorithmically Located Stressed "Syllables" with Perceived Stress Patterns, for the 6 ARPA Sentences.

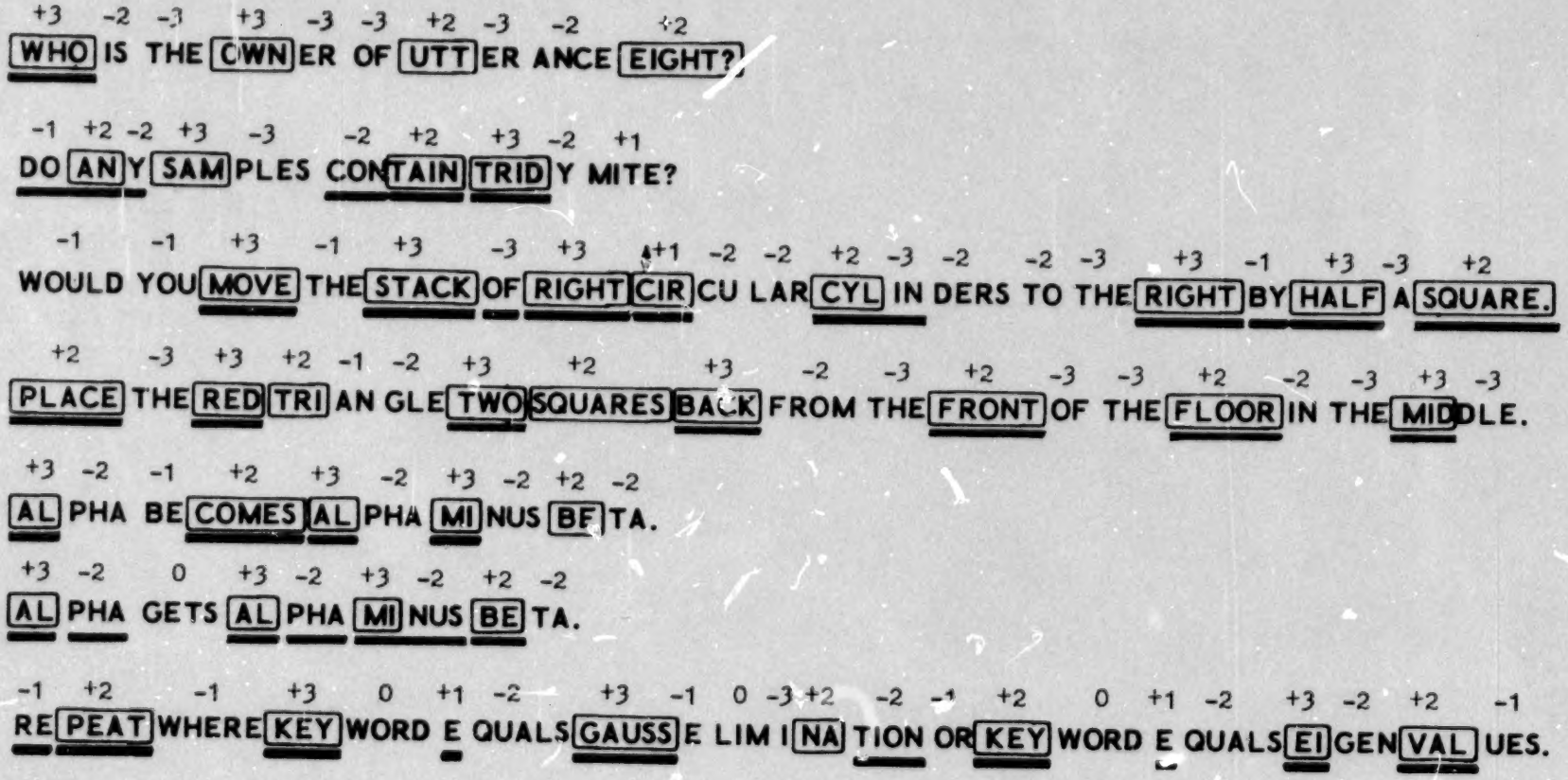

+3 -2 -3 +3 -3 -3 +2 -3 -2 +2
WHO IS THE OWN ER OF UTT ER ANCE EIGHT?
-1 +2 -2 +3 -3 -2 +2 +3 -2 +1
DO ANY SAMPLES CONTAIN TRIDY MITE?
-1 -1 +3 -1 +3 -3 +3 +1 -2 -2 +2 -3 -2 -2 -3 +3 -1 +3 -3 +2
WOULD YOU MOVE THE STACK OF RIGHT CIR CU LAR CYL IN DERS TO THE RIGHT BY HALF A SQUARE.
+2 -3 +3 +2 -1 -2 +3 +2 +3 -2 -3 +2 -3 -3 +2 -2 -3 +3 -3
PLACE THE RED TRI AN GLE TWO SQUARES BACK FROM THE FRONT OF THE FLOOR IN THE MIDDLE.
+3 -2 -1 +2 +3 -2 +3 -2 +2 -2
ALPHA BE COMES ALPHA MINUS BETA.
+3 -2 0 +3 -2 +3 -2 +2 -2
ALPHA GETS ALPHA MINUS BETA.
-1 +2 -1 +3 0 +1 -2 +3 -1 0 -3 +2 -2 -1 +2 0 +1 -2 +3 -2 +2 -1
REPEAT WHERE KEY WORD EQUALS GAUSS E LIM I NA TION OR KEY WORD EQUALS EI GEN VAL UES.

Figure C-11. Comparison of Algorithmically Located Stressed "Syllables" with Perceived Stress Patterns, for the 7 ARPA Sentences.